

Three papers on
JACKKNIFE STATISTICS

by

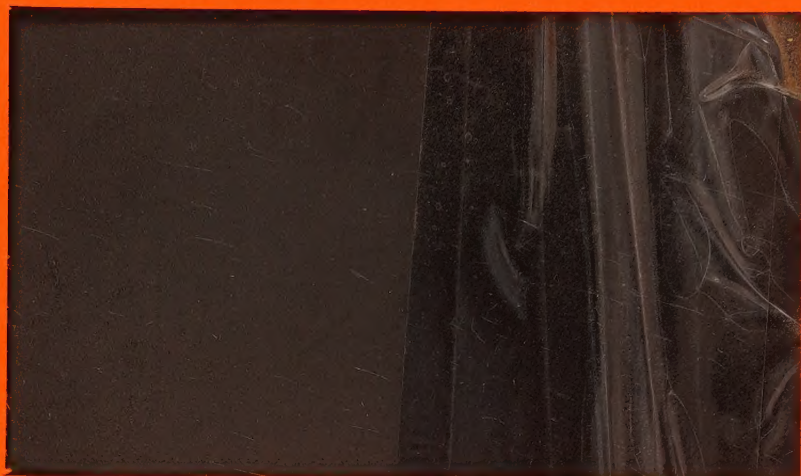
Daniel Thorburn



UNIVERSITY OF LUND AND
LUND INSTITUTE OF TECHNOLOGY
DEPARTMENT OF MATHEMATICAL STATISTICS

LUND, SWEDEN

1975



Three papers on
JACKKNIFE STATISTICS

by

Daniel Thorburn

SOME ASYMPTOTIC PROPERTIES OF
JACKKNIFE STATISTICS

by

Daniel Thorburn

1975:6

June 1975

SUMMARY

The jackknife and the original statistic often have the same asymptotic distribution. In fact, their difference often has a variance of smaller order than the original statistic. We give sufficient and almost necessary conditions for this. We prove a similar result also for regular functions of the statistic. Applications are given to sample moments and linear combinations of order statistics. We discuss possible gains and losses in precision when using the jackknife procedure. In particular we give a lower bound on the variance of the jackknife.

KEY WORDS

Asymptotic distribution, Jackknife, Moments, Order statistics, Variance.

SOME ASYMPTOTIC PROPERTIES OF JACKKNIFE STATISTICS

by

Daniel Thorburn

1. INTRODUCTION

Consider a sequence of independent and identically distributed random variables $X_1, X_2, \dots$ and real-valued symmetric functions g_n of n variables. For simplicity, let g_n also denote the random variable $g_n(X_1, \dots, X_n)$. Writing

$$g_{n,i} = g_{n-1}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n),$$

we define the jackknife statistic $\hat{g}_n$ by

$$(1.1) \quad \hat{g}_n = \frac{1}{n} \sum_{i=1}^n Y_{n,i}$$

where

$$(1.2) \quad Y_{n,i} = ng_n - (n-1)g_{n,i}.$$

The variables $Y_{n,i}$ are by some authors called pseudo-values. In the sequel we use the circumflex to indicate the corresponding jackknife statistic.

The jackknife statistic has often less bias than g_n . In fact a bias of $a \cdot n^{-k}$ is changed into a bias of order $a(1-k)(n^{-k} + kn^{-1-k}/2)$. When $k=1$ the first order term is completely removed. When k lies strictly between zero and two further improvement may be reached by jackknifing the jackknife statistic.

In this paper we shall not study the mean value properties of the jackknife as much as its asymptotic distribution. It is often normal with the same variance as the asymptotic one of the original statistic. We shall give a partial explanation of this fact.

We first give two conditions which guarantee that the asymptotic distribution of the jackknife is the same as that of the original statistic. These conditions give a more exact meaning to the idea of Tukey (1958) that the jackknife statistic can be regarded as a sum of n almost independent terms. In the third section we prove this fact and discuss the necessity of the conditions. In Section 4 we study functions of one or several statistics which obey the two conditions. In the fifth section we extend the results to

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41553215_0009

the more general case of exchangeable variables. In Section 6 we give applications to estimates obtained according to the moment method and to sums of order statistics.

In the last section we abandon the two conditions and discuss the asymptotic variance of the jackknife. In particular we show that if $V(g_n) \sim a/n$, the jackknife cannot be asymptotically more efficient than the original statistic g_n .

2. CONDITIONS

We give two conditions which together imply that $\hat{g}_n$ and g_n have the same asymptotic distribution. The proof of this fact will, however, not appear until the next section. In these conditions and in the sequel $o(\cdot)$ is taken as n tends to infinity.

CONDITION 2.1 The variance of the statistic g_n can be written

$$(2.1) \quad V(g_n) = \sigma^2/n + \epsilon_n$$

where

$$(2.2) \quad \epsilon_n = o(n^{-1}),$$

$$(2.3) \quad \epsilon_n - \epsilon_{n-1} = o(n^{-2}),$$

$$(2.4) \quad \epsilon_n - 2\epsilon_{n-1} + \epsilon_{n-2} = o(n^{-3}).$$

CONDITION 2.2 The conditional expectation of g_{n+1} given $X_1, \dots, X_n$ can be written as

$$(2.5) \quad E(g_{n+1} | X_1, \dots, X_n) = \frac{n}{n+1} g_n + r_n$$

where $r_n = r_n(X_1, \dots, X_n)$ is a random variable such that

$$(2.6) \quad V(r_n) = o(n^{-3}),$$

$$(2.7) \quad \text{Cov}(g_n, r_n) = \delta_n = o(n^{-2}),$$

$$(2.8) \quad \delta_n - \delta_{n-1} = o(n^{-3}).$$

The important formulas in these conditions are (2.1) and (2.5) together

with (2.2) and (2.6). The remaining four formulas are, as we shall see, regularity conditions.

Condition 2.1 holds if the variance of g_n can be written as

$$V(g_n) = \frac{\sigma^2}{n} + \frac{a}{n^2} + \frac{b}{n^3} + o(n^{-3}),$$

where a and b are arbitrary constants.

If ε_n can be written as a function, $f(1/n)$, with a bounded second derivative in a neighbourhood of zero, (2.3) and (2.4) can be proved from (2.1) and (2.2), as we shall now show. If f has a first derivative, it must be zero at zero since

$$(2.9) \quad f'(0) = \lim_{n \rightarrow \infty} n(f(0) - f(1/n)) = \lim_{n \rightarrow \infty} n(0 - o(n^{-1})) = 0.$$

We now get that

$$(2.10) \quad \varepsilon_n - \varepsilon_{n-1} = f(1/n) - f(1/(n-1)) = -1/(n^2-n)f'(\xi_n) = o(n^{-2}),$$

where ξ_n is a number between $1/n$ and $1/(n-1)$. We also find

$$\begin{aligned} (2.11) \quad \varepsilon_n - 2\varepsilon_{n-1} + \varepsilon_{n-2} &= f(1/n) - 2f(1/(n-1)) + f(1/(n-2)) = \\ &= -1/(n^2-n)f'(\xi_n) + 1/(n^2-3n+2)f'(\xi_{n-1}) = \\ &= 2/(n^3-3n^2+2n)f'(\xi_{n-1}) + 1/(n^2-n)(\xi_{n-1}-\xi_n)f''(\eta_n), \end{aligned}$$

where η_n is a number between ξ_n and ξ_{n-1} . The result (2.4) follows since $|\xi_n - \xi_{n-1}|$ is less than $1/(n-2) - 1/n$.

Condition 2.2 says that g_{n+1} is in some sense approximately a linear function of g_n and X_{n+1} . For instance, if we take g_n to be the arithmetic mean of n independent random variables, r_n is a constant, which does not depend on $X_1, \dots, X_n$.

It is easy to see that (2.7) may be proved from (2.1), (2.2) and (2.6). Condition (2.8) is, as we have already said, a regularity condition. For example, if (2.1), (2.2) and (2.6) holds and if δ_n can be written as a function of $1/n$, say $f(1/n)$, and if $f(x)/x$ has a continuous derivative at the origin, then (2.8) can be proved in the same way as (2.10).

3. THE JACKKNIFE

According to Tukey (1958) the variables

$$Y_{n,i} = n g_n - (n-1)g_{n,i}, \quad i = 1, 2, \dots, n,$$

can often to a good approximation be treated as n independent variables. We shall now prove that they are asymptotically uncorrelated if Conditions 2.1 and 2.2 hold. Note that the conditions do not necessarily imply asymptotic independence.

THEOREM 3.1 If Conditions 2.1 and 2.2 hold,

$$V(Y_{n,i}) = \sigma^2 + o(1)$$

$$\text{Cov}(Y_{n,i}, Y_{n,j}) = o(1/n), \quad i \neq j.$$

PROOF The proof consists of simple and straightforward computations. For the variance we use in turn the symmetry of g_n , the definitions of $Y_{n,i}$ and of r_n , the definitions (2.1) and (2.7) of ε_n and of δ_n and at the end formulas (2.2), (2.3) and (2.7).

$$\begin{aligned} V(Y_{n,i}) &= V(Y_{n,n}) = \\ &= V(n g_n - (n-1)g_{n-1}) = \\ &= n^2 V(g_n) + (n-1)^2 V(g_{n-1}) - 2n(n-1) \text{Cov}(g_n, g_{n-1}) = \\ &= n^2 V(g_n) + (n-1)^2 V(g_{n-1}) - 2n(n-1) ((n-1)/n V(g_{n-1}) + \text{Cov}(r_{n-1}, g_{n-1})) = \\ &= n^2 \left(\frac{\sigma^2}{n} + \varepsilon_n \right) - (n-1)^2 \left(\frac{\sigma^2}{n-1} + \varepsilon_{n-1} \right) - 2n(n-1) \delta_{n-1} = \\ &= \sigma^2 + n^2 (\varepsilon_n - \varepsilon_{n-1}) + (2n-1) \varepsilon_{n-1} - 2n(n-1) \delta_{n-1} = \\ &= \sigma^2 + o(1). \end{aligned}$$

Similarly we can compute the covariance. But first we observe that

$$\begin{aligned}
(3.2) \quad \text{Cov}(g_{n,1}, g_{n,2}) &= \text{Cov}(E(g_{n,1} | X_1, X_3, \dots, X_n), g_{n,2}) = \\
&= \text{Cov}(E(g_{n,1} | X_3, \dots, X_n), g_{n,2}) = \\
&= \text{Cov}(E(g_{n,1} | X_3, \dots, X_n), E(g_{n,2} | X_3, \dots, X_n)) = \\
&= \text{Cov}\left(\frac{n-2}{n-1} g_{n-2} + r_{n-2}, \frac{n-2}{n-1} g_{n-2} + r_{n-2}\right).
\end{aligned}$$

In the following calculations the same methods are used as in (3.1). In the last equality formulas (2.2), (2.3), (2.4), (2.6), (2.7) and (2.8) are used.

$$\begin{aligned}
\text{Cov}(Y_{n,i}, Y_{n,j}) &= \text{Cov}(ng_n^{-(n-1)} g_{n,1}, ng_n^{-(n-1)} g_{n,2}) = \\
&= n^2 V(g_n) - 2n(n-1) \text{Cov}(g_n, g_{n,1}) + (n-1)^2 \text{Cov}(g_{n,1}, g_{n,2}) = \\
&= n^2 V(g_n) - 2n(n-1) ((n-1)/n V(g_{n-1}) + \text{Cov}(r_{n-1}, g_{n-1})) + \\
&+ (n-1)^2 \text{Cov}\left(\frac{n-2}{n-1} g_{n-2} + r_{n-2}, \frac{n-2}{n-1} g_{n-2} + r_{n-2}\right) = \\
&= n^2 V(g_n) - 2(n-1)^2 V(g_{n-1}) - 2n(n-1) \text{Cov}(g_{n-1}, r_{n-1}) + \\
&+ (n-2)^2 V(g_{n-2}) + 2(n-1)(n-2) \text{Cov}(g_{n-2}, r_{n-2}) + V(r_{n-2}) = \\
&= n^2 \left(\frac{\sigma^2}{n} + \epsilon_n\right) - 2(n-1)^2 \left(\frac{\sigma^2}{n-1} + \epsilon_{n-1}\right) + (n-2)^2 \left(\frac{\sigma^2}{n-2} + \epsilon_{n-2}\right) - \\
&- 2n(n-1) \delta_{n-1} + 2(n-1)(n-2) \delta_{n-2} + (n-1)^2 V(r_{n-2}) = \\
&= n^2 (\epsilon_n - 2\epsilon_{n-1} + \epsilon_{n-2}) + (4n-2)(\epsilon_{n-1} - \epsilon_{n-2}) + 2\epsilon_{n-2} - \\
&- 2n(n-1)(\delta_{n-1} - \delta_{n-2}) - 4(n-1)\delta_{n-2} + (n-1)^2 V(r_{n-2}) = \\
&= o(n^{-1}). \quad \square
\end{aligned}$$

We can now compute the variance of the jackknife statistic. We formulate this as a corollary.

COROLLARY 3.2 If Conditions 2.1 and 2.2 hold,

$$V(\hat{g}_n) = \sigma^2/n + o(n^{-1}).$$

PROOF By the previous theorem and the definition (1.1)

$$\begin{aligned}
V(\hat{g}_n) &= V\left(\frac{1}{n} \sum_{i=1}^n Y_{n,i}\right) = \frac{1}{n^2} \sum_{i,j=1}^n \text{Cov}(Y_{n,i}, Y_{n,j}) = \\
&= n^{-2} \sum_{i=1}^n V(Y_{n,i}) + n^{-2} \sum_{i \neq j} \text{Cov}(Y_{n,i}, Y_{n,j}) = \\
&= n^{-2} (n\sigma^2 + o(1)) + n^{-2} n(n-1)o(n^{-1}) = \\
&= \sigma^2/n + o(n^{-1}).
\end{aligned}$$

□

A much sharper result is true. We can, in fact, show that $\hat{g}_n$ and g_n must have not only the same asymptotic variance but also the same asymptotic distribution, except possibly for a change of the location parameter. This is a consequence of the following theorem.

THEOREM 3.3 If Conditions 2.1 and 2.2 hold,

$$V(\hat{g}_n - g_n) = o(n^{-1}).$$

PROOF As in the previous proofs the computations are simple and straightforward.

We use the symmetry, the definition (2.5) of r_n , Corollary 3.2 and the definitions (2.1) and (2.7) of ε_n and δ_n and, for the last equality, formulas (2.2), (2.3) and (2.7).

$$\begin{aligned}
(3.3) \quad V(\hat{g}_n - g_n) &= \\
&= V(\hat{g}_n) + V(g_n) - 2\text{Cov}(g_n, n^{-1} \sum_{i=1}^n (ng_n - (n-1)g_{n,i})) = \\
&= V(\hat{g}_n) + V(g_n) - 2\text{Cov}(g_n, ng_n - (n-1)g_{n-1}) = \\
&= V(\hat{g}_n) + V(g_n) - 2nV(g_n) + 2(n-1)\left(\frac{n-1}{n} V(g_{n-1}) + \text{Cov}(r_{n-1}, g_{n-1})\right) = \\
&= \left(\frac{2}{n} - 2\right)\sigma^2 + o\left(\frac{1}{n}\right) - 2n\varepsilon_n + 2\frac{(n-1)^2}{n} \left(\frac{\sigma^2}{n-1} + \varepsilon_{n-1}\right) + 2(n-1)\delta_{n-1} = \\
&= 2n(\varepsilon_{n-1} - \varepsilon_n) - (4 - \frac{2}{n})\varepsilon_{n-1} + 2(n-1)\delta_{n-1} + o(n^{-1}) = \\
&= o(n^{-1}).
\end{aligned}$$

□

If the original statistic is asymptotically normal, this result is especially interesting. This is seen from the following corollary.

COROLLARY 3.4 If Conditions 2.1 and 2.2 hold and if

$$\sqrt{n}(g_n - E(g_n)) \xrightarrow{d} N(0, \sigma^2),$$

then

$$\sqrt{n}(\hat{g}_n - E(\hat{g}_n)) \xrightarrow{d} N(0, \sigma^2).$$

From the proofs of Theorems 3.1 and 3.3 it is clear that Conditions 2.1 and 2.2 are in some sense natural for those theorems. As an example of their necessity we give the following theorem:

THEOREM 3.5 Let $V(g_n) = \sigma^2/n + o(n^{-1})$ and suppose that this variance can be written as a function of $1/n$ with a bounded second derivative in a neighbourhood of zero. Suppose also that

$$\delta_n = k n^{-a} + o(n^{-a}), \quad k \neq 0, \quad a > 0,$$

where δ_n is defined by (2.5) and (2.7) and that $n^{a-1}\delta_n$ in a neighbourhood of zero can be written as a continuously differentiable function of $1/n$.

Then

$$V(\hat{g}_n - g_n) = o(n^{-1}) \Leftrightarrow \text{Condition 2.2.}$$

PROOF In Theorem 3.3 we proved the if-part. It remains to show only if. We proved above (formulas (2.10) and (2.11)) that Condition 2.1 must hold in this situation.

Using the same arguments as those leading to (3.3) we have

$$\begin{aligned} (3.4) \quad V(\hat{g}_n - g_n) &= \\ &= -2(n-1)^2(\delta_{n-1} - \delta_{n-2}) - 4 \frac{(n-1)^2}{n} \delta_{n-2} + \frac{(n-1)^3}{n} V(r_{n-2}) + o(n^{-1}). \end{aligned}$$

In the same way as we proved (2.10), we can prove that

$$(3.5) \quad \delta_{n-1} - \delta_{n-2} = -kan^{-1-a} + o(n^{-1-a}).$$

Let us consider three different situations depending on the value of a .

$a < 2$ In this situation the third term on the right hand side of (3.4) is the dominating one, since

$$n^2 V(r_n) \geq n^2 \delta_n^2 / (V(g_n)) = \frac{kn^{2-a}}{\sigma^2} \cdot n\delta_n + o(n^{3-2a})$$

and (3.5) holds. This means that $V(\hat{g}_n - g_n)$ is not $o(n^{-1})$. We have thus arrived at a contradiction.

$a = 2$ We can in this situation simplify (3.4):

$$\begin{aligned} V(\hat{g}_n - g_n) &= 2(n-1)^2 k a n^{-1-a} + o(n^{-1-a}) - 4(n-1)^2/n k n^{-a} + o(n^{1-a}) + \\ &\quad + (n-1)^3/n V(r_{n-2}) + o(n^{-1}) = \\ &= (n-1)^3/n V(r_{n-2}) + o(n^{-1}). \end{aligned}$$

From this it follows that $V(r_{n-2}) = o(n^{-3})$. This in turn implies that $\delta_n = \text{Cov}(g_n, r_n) = o(n^{-2})$, which is a contradiction.

$a > 2$ In this situation Condition 2.2 holds, since from (3.5) and (3.4) we have

$$V(\hat{g}_n - g_n) = \frac{(n-1)^3}{n} V(r_{n-2}) + o(n^{-1}).$$

This gives $V(r_{n-2}) = o(n^{-3})$. □

Conditions 2.1 and 2.2 are of course not necessary for convergence without any further assumptions. E.g. if g_n is the arithmetic mean of n Cauchy variables, the result of Theorem 3.3 is true but neither Condition 2.1 nor 2.2 holds. If we study the mean of normal variables, $\bar{X}$, and put $g_n = (\bar{X})^{-1}$, we have convergence in distribution if $E(\bar{X}) \neq 0$ but no moments exist.

4. JACKKNIFE OF A FUNCTION

Miller (1964) showed that if $(X_i)_{i=1}^{\infty}$ is a sequence of i.i.d. random variables with mean zero and positive standard deviation, σ , and if f is a function on the real line which in a neighbourhood of the origin has a bounded second derivative, then as n tends to infinity

$$\sqrt{n}(\widehat{f(\bar{X})} - f(0)) \xrightarrow{d} N(0, \sigma^2 (f'(0))^2).$$

With slight extensions of Miller's proof we have the following result in our situation.

THEOREM 4.1 Suppose that g_n satisfies Conditions 2.1 and 2.2 and that there exists a number m such that

$$(4.1) \quad \left. \begin{aligned} \sqrt{n}(E(g_n) - m) &\rightarrow 0 \\ \sqrt{n}(E(\hat{g}_n) - m) &\rightarrow 0 \end{aligned} \right\} \text{ as } n \rightarrow \infty.$$

Let f be a function, which in a neighbourhood of m has a bounded second derivative. In this situation we have

$$\sqrt{n}(\widehat{f(g_n)} - f(g_n)) \xrightarrow{d} 0 \text{ as } n \rightarrow \infty.$$

PROOF We can without loss of generality assume that $m = 0$.

Let d and M be positive quantities such that $|f''(x)| \leq M$ when $|x| \leq d$.

By the Tjebysjev inequality and Theorem 3.1 we have

$$(4.2) \quad P(|Y_{n,i} - E(Y_{n,i})|/n > d/3) \leq \frac{10\sigma^2}{n^2 d^2}$$

if n is large enough. From this we get

$$(4.3) \quad P(\max_i (|Y_{n,i} - E(Y_{n,i})|)/n > d/3) \leq \frac{10\sigma^2}{nd^2} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

We know that g_n tends to zero in probability as n tends to infinity. This fact, formula (4.3) and the equality

$$g_{n,i} - E(g_{n,i}) = \frac{n}{n-1} [g_n - E(g_n) - \frac{1}{n}(Y_{n,i} - E(Y_{n,i}))]$$

together imply that

$$P(-\frac{2}{3}d < g_n - E(g_n) < \frac{2}{3}d, -\frac{2}{3}d < g_{n,i} - E(g_{n,i}) < \frac{2}{3}d \text{ for } i \leq n) \rightarrow 1,$$

as n tends to infinity. We know by assumption that $E(g_n)$ tends to zero.

Hence we get

$$(4.4) \quad P(-d < g_n < d, -d < g_{n,i} < d \text{ for } i \leq n) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

In the probability arguments in the following we sometimes assume that g_n and all $g_{n,i}$ have absolute values less than d , and sometimes we do not make this assumption. By (4.4) the limit probabilities will be unaffected by this.

For g_n and $g_{n,i}$, with absolute values less than d , we have

$$(4.5) \quad f(g_{n,i}) = f(g_n) + (g_{n,i} - g_n) f'(g_n) + (g_{n,i} - g_n)^2 f''(\xi_i) / 2$$

where ξ_i is a number between $g_{n,i}$ and g_n . Hence we get

$$(4.6) \quad \widehat{f(g_n)} - f(g_n) = \frac{n-1}{n} \sum_1^n (f(g_n) - f(g_{n,i})) =$$

$$= - \frac{n-1}{n} f'(g_n) \sum_1^n (g_{n,i} - g_n) - \frac{n-1}{2n} \sum_1^n (g_{n,i} - g_n)^2 f''(\xi_i) =$$

$$= f'(g_n) (g_n^\wedge - g_n) - \frac{1}{2n(n-1)} \sum_1^n (g_n - Y_{n,i})^2 f''(\xi_i).$$

We shall estimate the two terms separately. First, by Theorem 3.3 and the assumption (4.1), we have

$$\sqrt{n}(g_n^\wedge - g_n) \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

Secondly we obtain

$$(4.7) \quad n^{-3/2} \left| \sum_1^n (g_n - Y_{n,i})^2 f''(\xi_i) \right| \leq n^{-3/2} M \sum_1^n (g_n - Y_{n,i})^2 =$$

$$= n^{-3/2} M \sum_1^n Y_{n,i}^2 - 2M\sqrt{n} g_n^\wedge \sum_1^n Y_{n,i} + M\sqrt{n} g_n^2.$$

The last two terms in this expression tend to zero in probability as n tends to infinity. By Theorem 3.1 and formula (4.1) we have that

$$E(n^{-1} \sum_1^n Y_{n,i}^2) = V(Y_{n,n}) + E^2(Y_{n,n}) = \sigma^2 + o(1) + o(n^{-1}).$$

The expression (4.7) thus tends to zero in probability. From (4.6) we now get

$$\sqrt{n}(\widehat{f(g_n)} - f(g_n)) \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty. \quad \square$$

It is necessary to make an assumption like (4.1). If we only assume that

$$\left. \begin{aligned} E(g_n) - m &\rightarrow 0 \\ E(g_n^\wedge) - m &\rightarrow 0 \end{aligned} \right\} \quad \text{as } n \rightarrow \infty$$

we only get

$$(4.8) \quad \sqrt{n}(\widehat{f(g_n)} - f(g_n)) - \sqrt{n} f'(E(g_n))(E(g_n^\wedge) - E(g_n)) \xrightarrow{d} 0.$$

This follows from formula (4.6) after some considerations. The expression (4.8) cannot be simplified. The assumptions (4.1) hold, however, in almost all situations. They say that the original bias is smaller than the standard deviation and that the expected value does not oscillate too wildly.

We cannot hope to prove that Conditions 2.1 and 2.2 hold for $f(g_n)$. We do not even know that the expected value or the variance exists. However, if we add some conditions on the function f and a condition concerning similarity of r_n and r_{n-1} , it is possible to show that $f(g_n)$ obeys Conditions 2.1 and 2.2. We will not do so here, since the computations are long and tedious and the result is not of any practical interest when we have Theorem 4.1.

In Theorem 4.1 we studied a fixed function of a one-dimensional statistic g_n . In many situations the function depends on n or g_n is a vector-valued random variable. With slight changes the proof of Theorem 4.1 holds in these situations too. First let us consider functions depending on n .

THEOREM 4.2 Let g_n satisfy Conditions 2.1 and 2.2 and let (4.1) hold. Let f_n be functions with first and second derivatives uniformly bounded in a neighbourhood of the origin. Suppose that

$$(4.9) \quad n^{3/2}(f_n(g_n) - f_{n-1}(g_n)) \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

Then

$$\sqrt{n}(\widehat{f_n(g_n)} - f_n(g_n)) \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

PROOF The proof is similar to the proof of Theorem 4.1. We shall therefore only sketch the proof.

Let d and M be positive quantities such that $|f'_n(x)| \leq M$ and $|f''_n(x)| \leq M$ for all $|x| \leq d$ and all n . In the same way as in the proof of Theorem 4.1, we conclude that we can assume that g_n and all $g_{n,i}$ have absolute values less than d , whenever we want to.

We have

$$(4.10) \quad \sqrt{n}(\widehat{f_n(g_n)} - f_n(g_n)) = \\ = \frac{n-1}{\sqrt{n}} \sum_1^n (f_{n-1}(g_n) - f_{n-1}(g_{n,i})) + (n-1)\sqrt{n}(f_n(g_n) - f_{n-1}(g_n)) =$$

$$= \sqrt{n} f'_{n-1}(g_n)(g_n - g_{n-1}) - \frac{1}{2\sqrt{n(n-1)}} \sum_1^n (g_n - g_{n,i})^2 f''_{n-1}(\xi_i) + \\ + \sqrt{n(n-1)}(f_n(g_n) - f_{n-1}(g_n)).$$

The rest of the proof is obvious. □

The next generalization is to a vector-valued statistic. Let g_n be a column vector, $(g_n)^T = (g_n^1, g_n^2, \dots, g_n^k)$, and let f be a function from R^k to R . We need some more notations. By the derivative of f we mean the row vector of partial derivatives, $f' = (f'_1, f'_2, \dots, f'_k)$, and by the second derivative the matrix of second partial derivatives,

$$f'' = \begin{pmatrix} f''_{11} & f''_{12} & \dots & f''_{1k} \\ f''_{21} & f''_{22} & \dots & f''_{2k} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ f''_{k1} & f''_{k2} & \dots & f''_{kk} \end{pmatrix}$$

By the absolute value of f' and of g_n we mean the Euclidean norm and by the absolute value of f'' the largest absolute value of the eigenvalues of f'' . This is well-defined since f'' is a symmetric matrix.

THEOREM 4.3 Let all elements of g_n satisfy Conditions 2.1 and 2.2 and (4.1). Let f be a real-valued and twice differentiable function with all second derivatives bounded in a neighbourhood of the origin. Then

$$\sqrt{n}(\widehat{f(g_n)} - f(g_n)) \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

PROOF The proof is similar to the proof of Theorem 4.1 and will only be sketched.

Let d and M be positive quantities such that $|f''(x)| \leq M$ when $|x| \leq d$. In the same way as in the proof of Theorem 4.1, we conclude that we can assume that g_n and all $g_{n,i}$ lie at a distance less than d from the origin. For such g_n and $g_{n,i}$ we have

$$f(g_{n,i}) = f(g_n) + f'(g_n)(g_{n,i} - g_n) + \frac{1}{2}(g_{n,i} - g_n)^T f''(\xi_i)(g_{n,i} - g_n)$$

where ξ_i is a column vector on the line between $g_{n,i}$ and g_n . Hence we get

$$\begin{aligned}
 (4.11) \quad \widehat{\sqrt{n}(f(g_n) - f(g_n))} &= \\
 &= \sqrt{n} f'(g_n)(\hat{g}_n - g_n) - \frac{1}{2\sqrt{n}(n-1)} \sum_1^n (g_n - Y_{n,i})^T f''(\xi_i)(g_n - Y_{n,i}).
 \end{aligned}$$

If we replace $f''(\xi_i)$ by the diagonal matrix with M in the diagonal, we make the absolute value of the second term larger. Using the theorem of Pythagoras, the second term can now be replaced by the corresponding sums for all the components of $(g_n - Y_{n,i})$:

$$M \sum_1^n |(g_n - Y_{n,i})|^2 = M \sum_1^k \sum_1^n (g_n^j - Y_{n,i}^j)^2$$

where $Y_{n,i}^j$ are the elements of $Y_{n,i}$. This and (4.7) used in (4.11), prove the theorem. $\square$

We can also combine the methods of Theorems 4.2 and 4.3.

COROLLARY 4.4 Let all the elements of g_n satisfy Conditions 2.1, 2.2 and (4.1).

Let $f_n: \mathbb{R}^k \rightarrow \mathbb{R}$ be twice differentiable functions with all first and second derivatives uniformly bounded in a neighbourhood of the origin. Suppose that

$$(4.12) \quad n^{3/2}(f_n(g_n) - f_{n-1}(g_n)) \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

Then

$$\widehat{\sqrt{n}(f_n(g_n) - f_n(g_n))} \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

In Section 3 we gave a theorem concerning the necessity of Conditions 2.1 and 2.2 (Theorem 3.5). Combining this theorem with the results of this section, we can from a knowledge of the behaviour of $\hat{g}_n$ say something about $\widehat{f(g_n)}$. A typical such theorem is the following one.

THEOREM 4.5 Suppose that $V(g_n) = \sigma^2/n + o(n^{-1})$, that f is a function and that the regularity conditions of Theorems 3.5 and 4.1 hold. If

$$\sqrt{n}(\hat{g}_n - g_n) \xrightarrow{m.s.} 0 \quad \text{as } n \rightarrow \infty,$$

then

$$\widehat{\sqrt{n}(f(g_n) - f(g_n))} \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

We have only studied one-dimensional functions of g_n . The generalization

to higher dimensions is immediate. From the convergence of each element, convergence of the whole vector follows.

5. EXTENSION TO EXCHANGEABLE VARIABLES

In this section we shall briefly show that, with some modifications, the results of the previous sections hold also for infinitely exchangeable variables.

When we computed $\text{Cov}(g_{n,1}, g_{n,2})$ in Theorem 3.1 we used the relation

$$(5.1) \quad E(g_{n-1} | X_1, X_2, \dots, X_{n-2}, X_n) = E(g_{n-1} | X_1, X_2, \dots, X_{n-2}),$$

which is automatically true for independent variables. If we make this equality into a condition, the results of Sections 3 and 4 will continue to hold for exchangeable variables with exactly the same proofs.

Another and perhaps more interesting way to study exchangeable variables is to work given the tail σ -algebra. If F_n is the σ -algebra generated by $X_n, X_{n+1}, \dots$, the tail σ -algebra is defined as the intersection of all these, $F = \bigcap_{n=1}^{\infty} F_n$. Now exchangeable variables are independent, given the tail σ -algebra, with probability one.

If we assume that, given the tail σ -algebra, Conditions 2.1 and 2.2 hold with probability one, we have that the results of Theorems 3.3 and 4.1 also hold with probability one. Now if a convergence in probability holds with probability one, it must also hold in probability unconditionally, and we have proved

THEOREM 5.1 If $X_1, X_2, \dots$ are exchangeable and if Conditions 2.1 and 2.2 hold given F with probability one, then

$$\sqrt{n}(g_n^\wedge - g_n) \xrightarrow{d} 0$$

and the results of Theorems 4.1, 4.2, 4.3 and Corollary 4.4 all hold.

6. SOME APPLICATIONS

To illustrate our results we shall study functions of moments and sums of order statistics. Some of the results are already known.

(a) Function of moments

Suppose that $X_1, X_2, \dots$ are i.i.d. random variables with finite moments of at least order $2k$. The arithmetic mean of X_i^k ,

$$\overline{X_i^k} = \frac{1}{n} \sum_{i=1}^n X_i^k,$$

satisfies Conditions 2.1 and 2.2 with ϵ_n and δ_n equal to zero.

If $f: \mathbb{R}^k \rightarrow \mathbb{R}$ is a fixed function that is twice differentiable with bounded second derivatives, it is well known that $f(\overline{X^1}, \overline{X^2}, \dots, \overline{X^k})$ has an asymptotic normal distribution. From Theorem 4.3 it follows that the corresponding jackknife statistic has the same property.

It is not hard to show directly that Conditions 2.1 and 2.2 hold for the sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} (\overline{X^2} - \bar{X}^2).$$

We shall, however, use Corollary 4.4 when studying functions of the sample variance. Let f have a finite second derivative and put $f_n(x, y) = f(n/(n-1) (y - x^2))$ and put x and y equal to the first two sample moments. Formula (4.12) holds for f_n defined in this way. We have thus that $f(s^2)$ has an asymptotic normal distribution.

PROPOSITION 6.1 If $X_1, X_2, \dots$ are i.i.d. random variables with finite fourth moment and mean zero

$$\sqrt{n}(\hat{s} - \sigma) \xrightarrow{d} N(0, (E(X^4) - \sigma^4)/(4\sigma^2)).$$

The jackknife technique has often been applied to ratio estimates.

Theorem 4.3 immediately gives many results in this situation, e.g. the following propositions, the proofs of which are obvious.

PROPOSITION 6.2 If $(X_1, Y_1), (X_2, Y_2), \dots$ are i.i.d. random pairs with finite second moments and with $E(Y_i) = m \neq 0$, and if f is a function with a bounded second derivative, then

$$\sqrt{n}(\widehat{f(\bar{X}/\bar{Y})} - f(\bar{X}/\bar{Y})) \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

PROPOSITION 6.3 If g'_n and g''_n both obey Conditions 2.1 and 2.2, and if (4.1) holds for g'_n and g''_n with $E(g''_n) \rightarrow m'' \neq 0$, then

$$\sqrt{n}(\widehat{(g'_n/g''_n)} - (g'_n/g''_n)) \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

Using Theorem 3.5 on this proposition we get

PROPOSITION 6.4 If g'_n and g''_n satisfy

$$V(\widehat{g'_n - g''_n}) = o(n^{-1}), \quad V(\widehat{g''_n - g''_n}) = o(n^{-1})$$

and the regularity conditions of Theorem 3.5, and if, also, (4.1) holds for g'_n and g''_n with $E(g''_n) \rightarrow m'' \neq 0$, then

$$\sqrt{n}(\widehat{(g'_n/g''_n)} - (g'_n/g''_n)) \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

(b) Linear combinations of order statistics

Suppose that $X_1, X_2, \dots$ are i.i.d. random variables with the distribution function F and that h is a function with a bounded second derivative on $(0, 1)$. We define

$$(6.1) \quad g_n = \frac{1}{n} \sum_{i=1}^n h\left(\frac{i-1}{n-1}\right) X_{(i,n)}$$

where $X_{(i,n)}$ is the i :th order statistic among $X_1, X_2, \dots, X_n$. It is well known that g_n , so defined, is asymptotically normal (Stigler 1974).

We shall show that $V(\widehat{g_n - g_n}) = o(n^{-1})$. It is often easier to verify this directly than to check Conditions 2.1 and 2.2. In view of Theorem 3.5 this is equivalent if some regularity conditions hold.

PROPOSITION 6.5 Let g_n be defined by (6.1). If h has a bounded second derivative and if $E(X^2)$ exists, then

$$\sqrt{n}(\widehat{g_n - g_n}) \xrightarrow{m.s.} 0 \quad \text{as } n \rightarrow \infty.$$

PROOF Using the mean value theorem twice, we have

$$\begin{aligned}
 (6.2) \quad \hat{g}_n &= \frac{1}{n} \sum_1^n \left(h\left(\frac{i-1}{n-1}\right) - (i-1)h\left(\frac{i-2}{n-2}\right) - (n-i)h\left(\frac{i-1}{n-2}\right) \right) X_{(i,n)} = \\
 &= \frac{1}{n} \sum_1^n \left(h\left(\frac{i-1}{n-1}\right) + \frac{(i-1)(n-i)}{(n-1)(n-2)} (h'(x_{1,i}) - h'(x_{2,i})) \right) X_{(i,n)} = \\
 &= g_n + \frac{1}{n} \sum_1^n \frac{(i-1)(n-i)}{(n-1)(n-2)} (x_{1,i} - x_{2,i}) h''(x_{3,i}) X_{(i,n)},
 \end{aligned}$$

where $x_{1,i}$ and $x_{2,i}$ are numbers between $(i-2)/(n-2)$ and $(i-1)/(n-1)$ and between $(i-1)/(n-1)$ and $(i-1)/(n-2)$, respectively, and where $x_{3,i}$ is between $x_{1,i}$ and $x_{2,i}$. Since $x_{2,i} - x_{1,i}$ is less than $1/(n-2)$ we get the result. $\square$

We have thus proved that the original and the jackknife statistic asymptotically have the same normal distribution. This also holds if in (6.1) we have $h(\frac{i}{n})$, $h(\frac{i+1}{n+1})$ or some similar expression. For such functions the difference between g_n and $\hat{g}_n$ also contains first derivatives of h .

With an example we now show that if h has a jump and if the distribution of X is continuous, then g_n does not obey Condition 2.2. Let X be uniformly distributed on the unit interval and let

$$h(x) = \begin{cases} 1 & \text{if } x \leq \frac{1}{2} \\ 0 & \text{if } x > \frac{1}{2} \end{cases}$$

Simple computations give

$$E(g_{2n} | X_1, \dots, X_{2n-1}) - \frac{2n-1}{2n} g_{2n-1} = -\frac{1}{4n} X_{(n,2n-1)}^2$$

$$E(g_{2n+1} | X_1, \dots, X_{2n}) - \frac{2n}{2n+1} g_{2n} = \frac{1}{2n+1} (X_{(n+1,2n)} - \frac{1}{2} X_{(n+1,2n)}^2).$$

The variances of these expressions are of order n^{-3} and Condition 2.2 can thus not hold.

If we have a quantile or a finite sum of quantiles, Condition 2.2 can of course not hold either. If h in (6.1) is continuous but not twice differentiable the situation is much more complicated. The result then depends on how badly h and the distribution F behave. We give one possible result which can be proved by calculations similar to the proof of Proposition 6.5.

If F is differentiable, h' has only a finite number of jumps and h'' is continuous and bounded elsewhere, then $\hat{g}_n - g_n$ tends to zero in mean square as n tends to infinity.

7. BOUNDS ON THE ASYMPTOTIC VARIANCE

In this section we first show that g_n can be regarded as an average over all possible jackknives for different n . Using this we obtain an upper bound for the variance of g_n expressed in terms of $V(\hat{g}_j)$, $j \leq n$. Finally we use this to get bounds for the relative efficiency of the jackknife statistic compared with g_n . In particular it is proved that if the variance of g_n is of order n^{-a} , where $a \leq 1$, then the asymptotic efficiency can not be higher than 1.

The jackknife is by (1.1) and (1.2) defined by

$$\hat{g}_n = ng_n - \frac{n-1}{n} \sum_1^n g_{n,i}.$$

Solving for g_n , we have

$$(7.1) \quad g_n = \frac{1}{n} \hat{g}_n + \frac{n-1}{n^2} \sum_1^n g_{n,i}.$$

From this equation it is seen that g_n can be chosen so that $\hat{g}_n$ is any desired function of $X_1, X_2, \dots, X_n$. This method of defining g_n can be repeated for all n . In particular we see that g_n can be chosen so that $\hat{g}_n$ is an efficient estimate, if such an estimate exists.

Using (7.1) on $g_{n,i}$ and inserting into (7.1) we get

$$g_n = \frac{1}{n} \hat{g}_n + \frac{1}{n} \bar{\hat{g}}_{n-1} + \frac{n-2}{n} \bar{g}_{n-2}$$

where $\bar{\hat{g}}_j$ is the arithmetic mean of all possible $\hat{g}_j(X_{i_1}, \dots, X_{i_j})$ and where $\bar{g}_{n-2}$ is the mean of all possible $g_{n-2}(X_{i_1}, \dots, X_{i_{n-2}})$.

Using this procedure iteratively we get the interesting formula

$$(7.2) \quad g_n = \frac{1}{n} \sum_1^n \bar{\hat{g}}_j.$$

This formula holds only when g_n is defined for all n . If g_n is defined only for $n \geq n_0$ for some n_0 , we can start the summation at n_0 defining $\hat{g}_{n_0} = g_{n_0}$, or in other words putting $g_n = 0$ for all $n < n_0$.

Formula (7.2) shows that g_n can be obtained from the jackknife by taking an average. It follows that the original statistic often behaves more smoothly than the jackknife. Note also that if for instance $\hat{g}_1 = g_1$ has a bias, then this bias will produce a bias of order $1/n$ in g_n .

Formula (7.2) can be used to estimate the variance of g_n in terms of $V(\hat{g}_j)$, $j = 1, 2, \dots, n$.

THEOREM 7.1 With the above mentioned assumptions we have

$$V(g_n) \leq \left(\frac{1}{n} \sum_1^n \left(\frac{j}{n} V(\hat{g}_j) \right)^{1/2} \right)^2.$$

PROOF Taking the variance of both sides of (7.2) we have

$$(7.3) \quad V(g_n) = n^{-2} V\left(\sum_1^n \bar{\hat{g}}_j\right) \leq n^{-2} \left(\sum_1^n (V(\bar{\hat{g}}_j))^{1/2}\right)^2.$$

The variable $\bar{\hat{g}}_j$ can be regarded as a U-statistic based on $\hat{g}_j$. In the theory of U-statistics (c.f. Fraser 1957, Section 6.5) it is proved that

$$V(\bar{\hat{g}}_j) \leq \frac{j}{n} V(\hat{g}_j).$$

Inserting this result into (7.3) we get the result. □

Theorem 7.1 looks somewhat simpler if its result is expressed in the standard deviation D instead of in the variance:

$$D(g_n) \leq \frac{1}{n} \sum_1^n (j/n)^{1/2} D(\hat{g}_j).$$

Cauchy's inequality and Theorem 7.1 give the following weaker version:

$$V(g_n) \leq \frac{1}{n} \sum_1^n V(\hat{g}_j).$$

Theorem 7.2 says that sometimes it is not possible to improve the asymptotic variance by means of the jackknife:

THEOREM 7.2 If $V(g_n) = C n^{-a} + o(n^{-a})$ for some positive $a \leq 1$, we have

$$(7.4) \quad \liminf_{n \rightarrow \infty} \frac{V(g_n)}{V(g_n^\wedge)} \leq 1.$$

PROOF The proof is by contradiction. If (7.4) does not hold, there must exist $C' < C$ and n_0 such that

$$(7.5) \quad V(g_n^\wedge) < C' n^{-a} \quad \text{for all } n > n_0.$$

By Theorem 7.1 this implies

$$\begin{aligned} V(g_n) &\leq \left(\frac{1}{n} \sum_{j=1}^n \left(\frac{j}{n} C' j^{-a} \right)^{1/2} + K n^{-3/2} \right)^2 \leq \\ &\leq n^{-3} \left(\left(\frac{2}{3-a} \right) C'^{1/2} n^{(3-a)/2} + K \right)^2 \leq \\ &\leq C' \left(\frac{2}{3-a} \right)^2 n^{-a} + o(n^{-a}) \leq \\ &\leq \frac{C'}{C} \left(\frac{2}{3-a} \right)^2 V(g_n) + o(V(g_n)) \end{aligned}$$

where $K n^{-3/2}$ is the contribution from those n where (7.5) does not hold. This expression is clearly a contradiction if $a \leq 1$. It follows that (7.4) must hold. □

If $V(g_n^\wedge) = C_2 n^{-a} + o(n^{-a})$ this theorem implies that

$$\lim_{n \rightarrow \infty} V(g_n)/V(g_n^\wedge) = C/C_2 \leq 1.$$

In fact we can by the same method as in the proof of Theorem 7.2 prove that if $V(g_n) = C n^{-a} + o(n^{-a})$, the following relation holds:

$$(7.6) \quad V(g_n) \leq \begin{cases} \left(\frac{2}{3-a} \right)^2 V(g_n^\wedge) + o(V(g_n^\wedge)) & \text{if } a < 3 \\ \ln(n) V(g_n^\wedge) + o(\ln(n) V(g_n^\wedge)) & \text{if } a = 3 \\ k n^{-3} & \text{if } a > 3 \end{cases}$$

where k is a constant.

Formula (7.6) implies that if a is strictly less than one, the variance of the jackknife is strictly larger than the original variance for large n . It also puts a limit on the possible improvement when a lies between one and three. The last cases are only included for completeness.

REFERENCES

- Fraser, D.A.S. (1957) Nonparametric methods in statistics. Wiley, New York.
- Miller, R.G. (1964) A trustworthy jackknife. Ann. Math. Statist. 35
pp 1594-1605.
- Miller, R.G. (1974) The jackknife - a review. Biometrika 61 pp 1-15.
- Tukey, J.W. (1958) Bias and confidence in not-quite large samples (Abstract)
Ann. Math. Statist. 29 p 614.

ON THE ASYMPTOTIC NORMALITY OF THE
JACKKNIFE

by

Daniel Thorburn

1975:7

June 1975

SUMMARY

The jackknife can be described as an arithmetic mean of the pseudo-values. If the pseudo-values converge in mean square to a sequence of random variables, these random variables are independent and identically distributed. This result is used to give conditions for the asymptotic normality of the jackknife. The variance estimator based on the pseudo-values is then consistent. The jackknife and the original statistic may have the same asymptotic distribution even if it is non-normal.

KEY WORDS

Asymptotic normality, Jackknife, Martingales, Pseudo-values, Variance estimator.

ON THE ASYMPTOTIC NORMALITY OF THE JACKKNIFE

by

Daniel Thorburn

1. INTRODUCTION

Let $X_1, X_2, \dots$ be a sequence of random variables, which are assumed to be exchangeable and in most sections also independent.

Let

$$g_n = g_n(X_1, \dots, X_n)$$

be real-valued symmetric functions of n random variables. Set

$$g_{n,i} = g_{n-1}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$$

$$Y_{n,i} = ng_n - (n-1)g_{n,i}.$$

The jackknife is defined as the arithmetic mean of the "pseudo-values"

$Y_{n,i}$,

$$\hat{g}_n = \frac{1}{n} \sum_{i=1}^n Y_{n,i}.$$

In this paper we always denote the jackknife by a circumflex.

In [1] the distributions of the jackknife and the original statistic are compared. It is shown that they are asymptotically equal under certain conditions. These conditions are given below. Here and in the sequel $o(\cdot)$ denotes the order when n tends to infinity.

CONDITION 1.1 $V(g_n) = \sigma^2/n + \epsilon_n$ where

$$(1.1) \quad \epsilon_n = o(n^{-1})$$

$$(1.2) \quad \epsilon_n - \epsilon_{n-1} = o(n^{-2})$$

$$(1.3) \quad \epsilon_n - 2\epsilon_{n-1} + \epsilon_{n-2} = o(n^{-3}).$$

CONDITION 1.2 $E(g_{n+1} | X_1, \dots, X_n) - \frac{n}{n+1} g_n = r_n$ where

$r_n = r_n(X_1, \dots, X_n)$ is a random variable such that

$$(1.4) \quad V(r_n) = o(n^{-3})$$

$$(1.5) \quad \text{Cov}(g_n, r_n) = \delta_n = o(n^{-2})$$

$$(1.6) \quad \delta_n - \delta_{n-1} = o(n^{-3}).$$

We shall reproduce some of the main results from [1]. The first one is the following theorem.

THEOREM 1.3 If Conditions 1.1 and 1.2 hold and if

$$(1.7) \quad E(g_{n-1} | X_1, \dots, X_{n-2}, X_n) = E(g_{n-1} | X_1, \dots, X_{n-2}),$$

then

$$(1.8) \quad V(Y_{n,i}) = \sigma^2 + o(1)$$

$$(1.9) \quad \text{Cov}(Y_{n,i}, Y_{n,j}) = o(n^{-1}) \quad i \neq j$$

$$(1.10) \quad V(\hat{g}_n - g_n) = o(n^{-1}).$$

Condition (1.7) holds for instance if $X_1, X_2, \dots$ are i.i.d. Another result in [1] deals with functions of statistics.

THEOREM 1.4 Let $X_1, X_2, \dots$ be i.i.d. and let Conditions 1.1 and 1.2 hold. Suppose that there exists a number m such that

$$(1.11) \quad \sqrt{n}(E(g_n) - m) \rightarrow 0 \quad \text{and} \quad \sqrt{n}(E(\hat{g}_n) - m) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Let f be a function which in a neighbourhood of m has a bounded second derivative. Then

$$\sqrt{n}(\widehat{f(g_n)} - f(g_n)) \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

The condition (1.11) holds trivially when

$$E(g_n) = m + k/n + o(n^{-3/2}).$$

This is thus a condition that holds in most situations where jackknifing is considered.

We finally give a theorem from [1] concerning the necessity of Condition 1.2:

THEOREM 1.5 Let $X_1, X_2, \dots$ be i.i.d. and let (1.1) hold. Suppose that ϵ_n can be written as a function of $1/n$ with a bounded second derivative, that

$$\delta_n = k n^{-a} + o(n^{-a}) \quad k \neq 0, a > 0,$$

and that $n^{a-1}\delta_n$ can be written as a continuously differentiable function of $1/n$. Then

$$V(\hat{g}_n - g_n) = o(n^{-1}) \Leftrightarrow \text{Condition 1.2.}$$

These are the results that we shall need from [1]. In this paper, we shall find out when the jackknife is asymptotically normal and we shall get estimates of its asymptotic variance.

We start in Section 2 by studying the variables $Y_{n,i}$. We also give a consistent estimate of the variance using these variables. Two different methods for proving asymptotic normality are then given in Section 3 and 4, respectively.

The first method, which is given in Section 3, builds on the results of Section 2. It is assumed that $X_1, X_2, \dots$ are independent. The jackknife is an arithmetic mean of the values $Y_{n,i}$. By the results (1.8), (1.9) in Theorem 1.3, these values are asymptotically uncorrelated with variance σ^2 . Hence it is tempting to believe that Conditions 1.1 and 1.2, which lead to this theorem, are sufficient for the asymptotic normality of the jackknife. However, this is not true and further conditions have to be introduced.

The second method is presented in Section 4. It utilizes the fact that the jackknife is asymptotically equal to the original statistic (Theorem 1.3). By means of martingale theory it is shown that g_n and hence also $\hat{g}_n$ is asymptotically normal.

Finally in Section 5 we discuss situations where the asymptotic distribution is not normal. We give two examples where Conditions 1.1 and 1.2 both hold.

2. THE PSEUDO-VALUES AND THE PSEUDO-VARIANCE

This section contains an investigation of the variables $Y_{n,i}$ and we assume that $X_1, \dots, X_n$ are i.i.d. Some of the results shall be used in the next section. Among other things we prove that if there exist variables Y_i such that all $Y_{n,i}$ converge to Y_i in mean square, these Y_i are independent. Using this we give limit results for a variance estimator.

The values $Y_{n,i}$ are sometimes called pseudo-values. This term indicates that they can be treated approximately as i.i.d. random variables. It is thus often possible to estimate the variance of g_n (and $\hat{g}_n$) by means of the "pseudo-variance":

$$\tilde{s}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_{n,i} - \hat{g}_n)^2.$$

When g_n is the arithmetic mean of a sequence of exchangeable variables such that Condition 1.1 holds, the pseudo-variance is not always a consistent estimate of σ^2 , defined in Condition 1.1. Nevertheless, many statisticians would in this situation prefer the pseudo-variance to the real variance, because $\hat{g}_n / \tilde{s}_n$ is often asymptotically normal. In Section 5 we discuss this example further.

In the remaining part of this section we assume that $X_1, X_2, \dots$ are i.i.d. The following condition is of fundamental importance for the work in this and the next section.

CONDITION 2.1 The pseudo-values $Y_{n,i}$ converge in mean square to random variables Y_i as n tends to infinity,

$$Y_{n,i} \xrightarrow{\text{m.s.}} Y_i \quad \text{as } n \rightarrow \infty \text{ for all } i.$$

Since $Y_{n,1}, \dots, Y_{n,n}$ are finitely exchangeable for each n , the variables $Y_1, Y_2, \dots$ are infinitely exchangeable. From Theorem 1.3 (formula (1.9)) it follows that they are also uncorrelated. The following theorem shows that they are even independent.

THEOREM 2.2 If $X_1, X_2, \dots$ are i.i.d. and if Condition 2.1 holds, then $Y_1, Y_2, \dots$ are independent and

$$E(Y_i | X_i) = Y_i.$$

Proof We shall first show that X_1 completely determines Y_1 . When X_1 is observed, we can always construct independent random variables $X'_2, X'_3, \dots$ with the same distribution as $X_2, X_3, \dots$. Given these constructed variables we obtain a limit variable $Y'_1 = Y_1(X_1, X'_2, X'_3, \dots)$. If we can show that Y'_1 does not depend on the values of $X'_2, X'_3, \dots$ we have proved the theorem for $i=1$. For this purpose we construct a new sequence $X''_2, X''_3, \dots$ and we shall show that $Y'_1 = Y''_1$.

For a fixed $\varepsilon > 0$ let n_0 be such that

$$E(Y_{n,1} - Y_1)^2 \leq \varepsilon \quad \forall n \geq n_0.$$

We get that

$$(2.1) \quad E(Y_{n,1} - Y_{m,1})^2 = E(Y_{n,1} - Y_1 - Y_{m,1} + Y_1)^2 < 4\varepsilon \quad \forall n, m \geq n_0.$$

We can now estimate the difference between Y'_1 and Y''_1 . First note that

$$\begin{aligned} & E(Y_{n,1}(X_1, X'_2, X'_3, \dots, X'_n) - Y_{n,1}(X_1, X''_2, X''_3, \dots, X''_n))^2 \rightarrow \\ & \rightarrow E(Y'_1 - Y''_1)^2 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

On the other hand this can by (2.1) be estimated by

$$\begin{aligned} & E(Y_{n,1}(X_1, X'_2, X'_3, \dots, X'_n) - Y_{n,1}(X_1, X''_2, X''_3, \dots, X''_n))^2 = \\ & = E(Y_{n,1}(X_1, X'_2, X'_3, \dots, X'_n) - Y_{2n-1,1}(X_1, X'_2, \dots, X'_n, X''_2, \dots, X''_n) + \\ & \quad + Y_{2n-1,1}(X_1, X''_2, \dots, X''_n, X'_2, \dots, X'_n) - Y_{n,1}(X_1, X''_2, X''_3, \dots, X''_n))^2 \leq \\ & \leq 16\varepsilon \quad \forall n \geq n_0. \end{aligned}$$

We now get that $E(Y'_1 - Y''_1)^2 = 0$. We have thus proved that we can compute Y_1 given X_1 , and this implies that $E(Y_1 | X_1) = Y_1$. From the symmetry it follows that $E(Y_i | X_i) = Y_i$. From the independence of the X -variables we have that the Y -variables are also independent. □

This theorem immediately yields several results for the Y_i .

Condition 2.1 gives that $V(Y_i) = \sigma^2$. The central limit theorem says that

$$(2.2) \quad \frac{1}{\sqrt{n}} \sum_1^n (Y_i - E(Y_i)) \xrightarrow{d} N(0, \sigma^2) \quad \text{as } n \rightarrow \infty.$$

The law of large numbers gives that

$$(2.3) \quad \frac{1}{n-1} (\sum_1^n Y_i^2 - \frac{1}{n} (\sum_1^n Y_i)^2) \xrightarrow{d} \sigma^2 \quad \text{as } n \rightarrow \infty.$$

We can now show that s_n is a consistent estimate of the standard deviation.

THEOREM 2.3 If $X_1, X_2, \dots$ are i.i.d. and if Conditions 1.1, 1.2 and 2.1 hold, then

$$s_n^2 \xrightarrow{d} \sigma^2 \quad \text{as } n \rightarrow \infty.$$

Proof We can without loss of generality assume that $E(Y_{n,i}) = E(Y_i) = 0$

The idea of the proof is to replace Y_i by $Y_{n,i}$ in (2.3). Using

Condition 2.1 we obtain

$$E\left(\left|\frac{1}{n} \sum_1^n (Y_i^2 - Y_{n,i}^2)\right|\right) \leq \frac{1}{n} \sum_1^n E(|Y_i^2 - Y_{n,i}^2|) = \frac{1}{n} n o(1) = o(1).$$

The independence of $Y_1, Y_2, \dots$ implies that

$$\frac{1}{n(n-1)} (\sum_1^n Y_i)^2 \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

Finally from Theorem 1.3 we get

$$E\left(\frac{1}{n(n-1)} (\sum_1^n Y_{n,i})^2\right) = \frac{1}{n(n-1)} \sum_1^n \sum_1^n \text{Cov}(Y_{n,i}, Y_{n,j}) = o(n^{-1}).$$

Inserting these results into (2.3) we obtain the desired result.

Theorem 2.3 continues to hold if we study a regular function of g_n :

THEOREM 2.4 Let $X_1, X_2, \dots$ be i.i.d. and suppose that g_n satisfies Conditions 1.1, 1.2, 2.1 and formula (1.11). Let f be a function with a bounded second derivative and $\tilde{s}_n$ be the pseudo-variance of $f(g_n)$. Then

$$\tilde{s}_n \xrightarrow{d} |f'(m)|\sigma \quad \text{as } n \rightarrow \infty.$$

Proof Without loss of generality we assume that $m = f(m) = 0$. Using the definition of $\tilde{s}_n^2$ and the mean value theorem a number of times, we obtain

$$\begin{aligned}
 (2.4) \quad \tilde{s}_n^2 + \frac{n}{n-1} \widehat{f(g_n)}^2 &= \frac{1}{n-1} \sum_1^n (nf(g_n) - (n-1)f(g_{n,i}))^2 = \\
 &= \frac{1}{n-1} \sum_1^n (nf(g_n) - (n-1)f(g_n) + (n-1)f'(g_n)(g_n - g_{n,i}) - \\
 &\quad - \frac{n-1}{2} f''(\xi_i)(g_n - g_{n,i})^2)^2 = \\
 &= \frac{1}{n-1} \sum_1^n (f'(0)g_n + f''(\eta_i)g_n^2/2 + (n-1)f'(0)(g_n - g_{n,i}) + \\
 &\quad + (n-1)f''(\rho_i)(g_n - g_{n,i})g_n - \frac{n-1}{2} f''(\xi_i)(g_n - g_{n,i})^2)^2 = \\
 &= \frac{1}{n-1} \sum_1^n (f'(0)(ng_n - (n-1)g_{n,i}) + f''(\eta_i)g_n^2/2 + (n-1) \cdot \\
 &\quad \cdot f''(\rho_i)(g_n - g_{n,i})g_n - \frac{n-1}{2} f''(\xi_i)(g_n - g_{n,i})^2)^2
 \end{aligned}$$

where ξ_i is a number between g_n and $g_{n,i}$ and where ρ_i and η_i are between 0 and g_n . Now since $n(g_n - g_{n,i})g_n \xrightarrow{m.s.} 0$ and since

$$\tilde{s}_n^2 + \frac{n}{n-1} \widehat{g_n}^2 = \frac{1}{n-1} \sum_1^n (ng_n - (n-1)g_{n,i})^2$$

we get:

$$\tilde{s}_n^2 + \frac{n}{n-1} \widehat{f(g_n)}^2 - (f'(0))^2 (\tilde{s}_n^2 + \frac{n}{n-1} \widehat{g_n}^2) \xrightarrow{d} 0 \text{ as } n \rightarrow \infty.$$

But by assumption (1.11) and Theorem 1.4

$$\widehat{g_n} \xrightarrow{d} 0 \quad \text{and} \quad f(g_n) \xrightarrow{d} 0.$$

Theorem 2.3 now gives the result.

The condition in this theorem that f'' is bounded everywhere is unnecessarily strong. It is sufficient to assume that f'' is bounded in a neighbourhood of m . This can be proved by showing that the probability that g_n and $g_{n,i}$ lie in this neighbourhood tends to one.

3. FIRST METHOD USING THE PSEUDO-VARIABLES

We shall in this section present our first method for proving asymptotic normality of the jackknife $\hat{g}_n$. For this purpose we shall use the pseudo-variables discussed in the previous section.

In that section we gave a limit result for sums of Y_i (2.2). We shall now replace Y_i in this expression by $Y_{n,i}$. To do so we need the following lemma, the result of which is quite natural in view of formula (1.9). Our proof is, however, not so simple.

LEMMA 3.1 If $X_1, X_2, \dots$ are i.i.d. and if Conditions 1.1, 1.2 and 2.1 hold, then

$$\text{Cov}(Y_{n,i}, Y_{n,j}) = o(n^{-1}), \quad i \neq j.$$

Proof We project $Y_{n,i}$ on the space spanned by $Y_1, \dots, Y_n$,

$$(3.1) \quad Y_{n,i} = a_n Y_i + b_n \bar{Y} + c_n Z_{n,i}$$

where a_n, b_n and c_n are real numbers, $\bar{Y} = \sum_1^n Y_j/n$, and $Z_{n,i}$ are residuals with variance one. Condition 1.2 says that

$$(3.2) \quad V(E(Y_{n,i} | X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)) = o(n^{-1}).$$

The expected value in this expression is by (3.1) and Theorem 2.2

$$(3.3) \quad b_n \frac{1}{n} \sum_{j \neq i} Y_j + E(c_n Z_{n,i} | X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n).$$

Since the first term here is uncorrelated with $Z_{n,i}$ and the conditioning includes $\sum_{j \neq i} Y_j$, the two terms are still uncorrelated. Formulas

(3.2) and (3.3) now give

$$b_n^2 \frac{n-1}{n^2} V(Y_j) = o(n^{-1}).$$

This means that b_n tends to Zero as n tends to infinity. The proof can now be finished,

$$\begin{aligned} \text{Cov}(Y_{n,i}, Y_{n,j}) &= \text{Cov}(a_n Y_i + b_n \bar{Y} + c_n Z_{n,i}, Y_j) = \\ &= b_n/n V(Y_j) = o(n^{-1}). \end{aligned}$$

□

We can now prove a limit theorem for $\hat{g}_n$:

THEOREM 3.2 If $X_1, X_2, \dots$ are i.i.d. and if Conditions 1.1, 1.2 and 2.1 hold, then

$$(3.4) \quad \sqrt{n}(\hat{g}_n - E(\hat{g}_n)) \xrightarrow{d} N(0, \sigma^2) \quad \text{as } n \rightarrow \infty$$

and

$$(3.5) \quad \sqrt{n}(\hat{g}_n - E(\hat{g}_n))/\tilde{s}_n \xrightarrow{d} N(0, 1) \quad \text{as } n \rightarrow \infty.$$

Proof We can without loss of generality assume that $E(Y_{n,i}) = E(Y_i) = 0$.

If Y_i could be replaced by $Y_{n,i}$ in formula (2.2), the first result would follow. But this is possible, since

$$\begin{aligned} (3.6) \quad V(\sqrt{n} \hat{g}_n - \frac{1}{\sqrt{n}} \sum_1^n Y_i) &= V(\frac{1}{\sqrt{n}} \sum_1^n (Y_{n,i} - Y_i)) \\ &= \frac{1}{n} (\sum_1^n V(Y_{n,i} - Y_i) + \sum_{i \neq j} (\text{Cov}(Y_{n,i}, Y_{n,j}) + \\ &\quad + \text{Cov}(Y_i, Y_j) - 2\text{Cov}(Y_{n,i}, Y_j))) = \\ &= o(1). \end{aligned}$$

For the last equality we have used Condition 2.1, Theorems 1.3, 2.2 and Lemma 3.1.

The result (3.5) is a simple consequence of Theorem 2.3 and the first result (3.4). □

It is interesting to note that by Theorem 1.3, the results of Theorem 3.2 also hold for the original statistic g_n . When g_n is an estimate of an unknown parameter, this theorem can be used to obtain good confidence intervals.

Combining Theorems 1.4 and 3.2 we obtain the following result.

THEOREM 3.3 Let $X_1, X_2, \dots$ be i.i.d. and suppose that Conditions 1.1, 1.2, 2.1 and formula (1.11) hold. If f has a bounded second derivative in a neighbourhood of m , then

$$\sqrt{n} (\widehat{f(g_n)} - f(m)) \xrightarrow{d} N(0, (f'(m)\sigma)^2) \quad \text{as } n \rightarrow \infty.$$

Proof The left hand side can be written

$$\begin{aligned} \sqrt{n} ((\widehat{f}(g_n) - f(g_n)) + (f(g_n) - f(m) - f'(m)(g_n - m)) + \\ + f'(m)(g_n - \hat{g}_n) + f'(m)(\hat{g}_n - m)). \end{aligned}$$

The first three terms all tend to zero in distribution (Theorem 1.4, standard asymptotic theory and Theorem 1.3). By Theorem 3.2 it follows that the last term is asymptotically normal.

By Theorem 2.4 the standard deviation $|f'(m)|\sigma$ can be estimated. We have thus that the results of Theorem 3.2 hold when g_n can be written as a function of a statistic such that the conditions hold.

If the regularity conditions of Theorem 1.5 and Condition 2.1 hold, we easily get the following result

$$nV(\hat{g}_n - g_n) \rightarrow 0 \quad \Rightarrow \quad \begin{cases} n(\hat{g}_n - m) \xrightarrow{d} N(0, \sigma^2) \\ \hat{s}_n \xrightarrow{d} \sigma. \end{cases}$$

The results of this section may be extended to exchangeable variables. If it is assumed that Conditions 1.1 and 1.2 hold given the tail σ -algebra (c.f. [1]) and if Condition 3.1 holds it is rather easy to show that $\hat{g}_n/\hat{s}_n$ is asymptotically normal unconditionally. The same result may be proved if Conditions 1.1, 1.2 and the following strengthened version of (1.7) holds:

$$E(g_{n-1} | X_1, \dots, X_{n-2}, F) = E(g_{n-1} | X_1, \dots, X_{n-2})$$

where F is the tail σ -algebra.

4. SECOND METHOD USING MARTINGALES

In this section we shall first show that ng_n is almost a martingale, if Condition 1.2 holds. Applying martingale theory we then obtain conditions under which g_n and hence also $\hat{g}_n$ is asymptotically normal.

We assume that the expected value of g_n is zero, which does not mean any loss of generality.

From the definition of r_n in Condition 1.1 we have

$$\begin{aligned} E(ng_n - (n-1)g_{n-1} | g_1, \dots, g_{n-1}) &= n E(r_{n-1} | g_1, \dots, g_{n-1}) = \\ &= n r'_{n-1}, \end{aligned}$$

say. Hence

$$(4.1) \quad ng_n - \sum_1^n i r'_{i-1}$$

is a martingale. By the definition of $Y_{n,i}$ we have

$$(4.2) \quad ng_n - \sum_1^n i r'_{i-1} = \sum_1^n (Y_{ii} - i r'_{i-1}).$$

Since (4.1) is a martingale, the right hand side of (4.2) is also a martingale.

We shall now prove that the remainder term $\sum_1^n i r'_{i-1}$ is negligible compared with $\sqrt{n} g_n$. Formula (1.4) says that for any given $\varepsilon > 0$ there exists an n_0 such that

$$(4.3) \quad n^3 V(r_{n-1}) < \varepsilon \quad \text{for all } n > n_0.$$

Since $r'_{n-1} = E(r_n | g_1, \dots, g_{n-1})$, we have that $V(r'_n) \leq V(r_n)$. Formula (4.3) thus holds also for r'_{n-1} . Using this on the remainder term, we find

$$\begin{aligned} E\left(\left|\frac{1}{\sqrt{n}} \sum_1^n i r'_{i-1}\right|\right) &\leq \frac{1}{\sqrt{n}} \sum_1^n i E(|r'_{i-1}|) \leq \\ &\leq \frac{1}{\sqrt{n}} \sum_1^{n_0-1} i E(|r'_{i-1}|) + \frac{1}{\sqrt{n}} \sum_{n_0}^n i \sqrt{V(r'_{i-1})} \leq \\ &\leq \frac{c}{\sqrt{n}} + \frac{\sqrt{\varepsilon}}{\sqrt{n}} \sum_{n_0}^n 1/\sqrt{i} \leq \\ &\leq \varepsilon + \frac{\sqrt{\varepsilon}}{\sqrt{n}} 2\sqrt{n} \end{aligned}$$

if n is large enough. Here c denotes a constant independent of n .

Using this in (4.2) we get

$$\frac{1}{\sqrt{n}} (ng_n - \sum_1^n (Y_{i,i} - ir_{i-1}')) \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

We can thus use limit theorems for martingales and the result will carry over to ng_n . There exist many central limit theorems for martingales. The following one is due to Scott (1973):

Let $Z_1, Z_2, \dots$ be martingale differences and let $s_n^2 = \sum_1^n V(Z_i)$.

If

$$(4.4) \quad \frac{1}{s_n^2} \sum_1^n Z_i^2 \xrightarrow{d} 1$$

$$(4.5) \quad \frac{1}{s_n^2} \sum_1^n Z_i^2 I(Z_i > \varepsilon s_n) \xrightarrow{d} 0, \quad \forall \varepsilon > 0$$

then

$$\frac{1}{s_n} \sum_1^n Z_i \xrightarrow{d} N(0,1).$$

When this theorem is applied to our situation, we get the following result:

THEOREM 4.1 Let formulas (1.1), (1.2) and (1.4) hold. If the distributions of $Y_{n,n}$ have bounded fourth moments and if they converge to distributions with variance σ^2 and if for $i, j > n$

$$(4.6) \quad \text{Cov}(Y_{i,i}^2, Y_{j,j}^2) = o(n^{-1}),$$

then

$$\sqrt{n} g_n \xrightarrow{d} N(0, \sigma^2) \quad \text{as } n \rightarrow \infty.$$

Proof We note that by formulas (1.1), (1.2) and (1.4) the variance of $Y_{n,n}$

is

$$\begin{aligned} V(Y_{n,n}) &= n^2 V(g_n) + (n-1)^2 V(g_{n-1}) - 2n(n-1) \text{Cov}(g_n, g_{n-1}) = \\ &= n^2 V(g_n) - (n-1)^2 V(g_{n-1}) - 2n(n-1) \delta_{n-1} = \\ &= \sigma^2 + n^2 \varepsilon_n - (n-1)^2 \varepsilon_{n-1} - 2n(n-1) \delta_{n-1} = \\ &= \sigma^2 + o(1). \end{aligned}$$

From this we get that $s_n^2 / (n\sigma^2) \rightarrow 1$. We can thus use $n\sigma^2$ instead of s_n^2 . We also note that

$$(4.7) \quad \frac{1}{n\sigma^2} \sum_1^n (Y_{i,i} - ir'_{i-1})^2 = \\ = \frac{1}{\sigma^2} \left(\frac{1}{n} \sum_1^n Y_{i,i}^2 - \frac{2}{n} \sum_1^n Y_{i,i} ir'_{i-1} + \frac{1}{n} \sum_1^n i^2 r_{i-1}'^2 \right).$$

From calculations similar to those in (4.3) we find that the last two terms tend to zero in distribution. In order to check the asymptotic normality we ought to check (4.4) and (4.5) for $Z_i = Y_{i,i} - ir'_{i-1}$, but by (4.7) it is equivalent to proving them for $Y_{i,i}$.

This is fairly simple, since the conditions are chosen so that (4.4) and (4.5) hold:

$$(4.8) \quad V\left(\frac{1}{n} \sum_1^n Y_{i,i}^2\right) = n^{-2} \left(\sum_1^n V(Y_{i,i}^2) + \sum_{j < i} 2\text{Cov}(Y_{i,i}^2, Y_{j,j}^2) \right).$$

Condition (4.6) and the assumption on bounded fourth moments imply that this expression tends to zero in distribution and (4.4) follows. Condition (4.5) is a simple consequence of the convergence of the distribution of $Y_{n,n}$ to a distribution with variance σ^2 . □

From (4.8) we see that $\frac{1}{n} \sum_1^n Y_{i,i}^2$ is a consistent estimate of the variance. If we strengthen Condition (4.6) a little to

$$\text{Cov}(Y_{i,m}^2, Y_{1,n}^2) = o(n^{-1}) \quad \text{for } m \geq n \quad \text{and } i > 1,$$

the pseudo-variance is also consistent. This may be proved exactly as (4.8).

Using the theorem by Scott, other and stronger results than Theorem 4.1 may be proved. It is interesting to note that both in Section 3 and 4 the extra condition deals with the convergence of the pseudo-values. Here in Theorem 4.1 we only assumed convergence in distribution but added some conditions on the moments, while in Section 3 the convergence was in mean square.

5. EXAMPLES OF NON-NORMALITY

In the previous sections we have given conditions for asymptotic normality. In this section we give examples where the asymptotic distribution is non-normal or does not exist. We first show that the jackknife can have any distribution if there are no conditions on the behaviour of g_n . After that we give two examples of asymptotic non-normality, where Conditions 1.1 and 1.2 hold.

Rewriting the definition of the jackknife we have

$$g_n = \frac{1}{n} g_n^{\wedge} + \frac{n-1}{n} \sum_1^n g_{n,i}.$$

This formula can be used to define g_n recursively so that $g_n^{\wedge}$ gets any desired distribution. If Conditions 1.1 and 1.2 hold for g_n this cannot be done.

We now give two examples where Conditions 1.1 and 1.2 hold, but where the asymptotic distribution is not normal.

EXAMPLE 5.1 This example is already mentioned at the beginning of section 2. It is not really a counter-example as it at the same time is an example of the situation described at the end of Section 3.

Let $Y_0, Y_1, Y_2, \dots$ be i.i.d. random variables which take the values plus and minus one with equal probability. Define

$$X_i = (Y_0 + 2)Y_i, \quad i = 1, 2, \dots$$

These variables have an exchangeable distribution. Taking g_n equal to the arithmetic mean, we have

$$V(g_n) = 5/n$$

$$E(g_{n+1} | X_1, \dots, X_n) = \frac{n}{n+1} g_n.$$

Conditions 1.1, 1.2 and (1.7) thus hold, but the asymptotic distribution is a mixture of two normal distributions with different standard deviations.

The tail σ -algebra is in this situation the algebra generated by Y_0 . Given Y_0 , the variable g_n is just the arithmetic mean of n independent variables. We easily see that Conditions 1.1, 1.2 and 3.1 hold given the tail σ -algebra and that

$$\sqrt{n} \hat{g}_n / \tilde{s}_n \xrightarrow{d} N(0,1)$$

holds unconditionally; a result that was predicated at the end of Section 3.

In the same way we can construct any mixture of normal distributions just by changing the distribution of Y_0 . If the variance of Y_0 exists Conditions 1.1, 1.2 and (1.7) also hold unconditionally.

In Example 5.1 the crucial fact is that $X_1, X_2, \dots$ are only exchangeable. But counter-examples may also be found when they are independent.

EXAMPLE 5.2 This is a very technical example but the idea is simple.

The variable g_n is the arithmetic mean of n independent variables, $X_{i,n}$, whose distribution depends on n . By letting $X_{i,n}$ be more and more skew we see to it that the limiting distribution is not normal. The problem is that if the distribution of $X_{i,n}$ changes too fast, Condition 1.2 will not hold, but if the change is too slow the skewness does not increase fast enough.

Let $n_0, n_1, n_2, \dots$ be integers such that $n_0 = 1$ and

$$(5.1) \quad \sum_{k=n_{j-1}+1}^{n_j-1} \frac{1}{k \ln(k)} < 1 \leq \sum_{k=n_{j-1}+1}^{n_j} \frac{1}{k \ln(k)}.$$

Define independent random variables $Z_{1,0}, Z_{1,1}, \dots$ such that

$$(5.2) \quad P(Z_{1,j} = x) = \begin{cases} \frac{1}{2} \exp(-2\exp(n_j)) & \text{if } |x| = \exp(\exp(n_j)) \\ 1 - \exp(-2\exp(n_j)) & \text{if } x = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Construct random variables which slowly shift from $Z_{1,0}$ to $Z_{1,1}$, from $Z_{1,1}$ to $Z_{1,2}$ and so on as n increases:

$$(5.3) \quad X_{1,n} = \cos(\pi K/2) Z_{1,j-1} + \sin(\pi K/2) Z_{1,j}$$

where K and j are numbers that depend on n and are defined by

$$(5.4) \quad n_{j-1} < n \leq n_j,$$

$$K = \min\left(\frac{n}{\sum_{k=n_{j-1}+1}^n \frac{1}{k \ln(k)}}, 1\right).$$

The next step is to define independent sequences $(X_{2,n})_1^\infty$, $(X_{3,n})_1^\infty$, with the same distribution as $(X_{1,n})_1^\infty$. Finally define the variable g_n as the arithmetic mean

$$g_n = \frac{1}{n} \sum_{i=1}^n X_{i,n}.$$

We see that g_n is a symmetric function of the sequences X_n . It remains to check that Conditions 1.1 and 1.2 hold but that the limiting distribution is not normal with variance one.

Condition 1.2 is a little harder to check. We have that

$$V(g_n) = n^{-2} \sum_{i=1}^n V(X_{i,n}) = n^{-2} n (\cos^2(\pi K/2) + \sin^2(\pi K/2)) = n^{-1}.$$

Condition 1.2 is a little harder to check. We have that

$$E(g_{n+1} | X_1, X_2, \dots, X_n) - \frac{n}{n+1} g_n = \frac{1}{n+1} \sum_{i=1}^n (X_{i,n+1} - X_{i,n}).$$

Some calculations will show that

$$V(r_n) = \frac{n}{(n+1)^2} 4 \sin^2[\pi/(4(n+1) \ln(n+1))] = o(n^{-3})$$

$$\text{Cov}(g_n, r_n) = \frac{1}{n+1} 2 \sin^2[\pi/(4(n+1) \ln(n+1))] = o(n^{-3}).$$

It is here assumed that $n+1 \neq n_j$ for any j . If $n+1 = n_j$ for some j the terms are even smaller. It is now easy to verify Condition 1.2.

It is easy to check that the limiting distribution cannot be normal with the right variance:

$$(5.5) \quad P(g_{n_j} \neq 0) \leq n_j P(Z_{1,j} \neq 0) = n_j \exp(-2 \exp(n_j)) \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

That X_n is defined to be a sequence and not a random variable is not important, since there exists a one to one correspondence between possible values of the sequence and the unit interval. Formula (5.5) can be true without the choice of the double exponential in (5.2); the value n_j would do equally well. We made this choice to indicate that $n_{i+1} \approx \exp(\exp(n_i))$. The expression $1/(k \ln(k))$ in (5.1) is more crucial. As a function of k , it must be $o(k^{-1})$, but its sum must diverge.

In Example 5.2 we can choose the distribution of $Z_{1,0}, Z_{1,1}, \dots$ arbitrarily. As long as they are independent with variance one, Conditions 1.1 and 1.2 continue to hold. If the skewness is large for every second $Z_{1,i}$ and small for the other, the distribution of $\sqrt{n} g_n$ does not even converge. It is also possible to choose the distributions so that g_n is asymptotically distributed according to the Poisson law along a subsequence of n .

6. AN APPLICATION TO THE RANK SUM

In this section we show that Conditions 1.1, 1.2 and 2.1 hold for a linear function of the Wilcoxon rank sum. Let $X_1, X_2, \dots$ and $Z_1, Z_2, \dots$ be two independent sequences of i.i.d. random variables with distribution functions F_X and F_Z , respectively, and such that $0 < P(X_1 \leq Z_1) < 1$. Define

$$g_n = \#\{(i,j); X_i \leq Z_j; i,j \leq n\}/n^2 = \sum_1^n F_{n,x}(Z_i)/n$$

where $\#$ denotes the number of elements in the set and where $F_{n,x}$ and $F_{n,z}$ denotes the empirical distribution functions of the first n X -variables and Y -variables, respectively. The Wilcoxon rank sum is a linear function of g_n , $n^2 g_n + n(n+1)/2$.

It is well known that Condition 1.1 holds. To show Condition (2.1) we get, after simple calculations, that

$$\begin{aligned}
 Y_{n,1} &= F_{n,x}(Z_1) + 1 - F_{n,z}(X_1) - I(X_1 \leq Z_1)/n - \\
 &\quad - \frac{1}{n(n-1)} \{(i,j); X_i \leq Z_j; 2 \leq i,j \leq n\} \xrightarrow{\text{m.s.}} \\
 &\xrightarrow{\text{m.s.}} F_x(Z_1) + 1 - F_z(X_1) - P(X \leq Z).
 \end{aligned}$$

Condition 1.2 is more difficult to show. We easily get that

$$\begin{aligned}
 r_n &= E(Y_{n+1,n+1} | X_1, X_2, \dots, Z_1, Z_2, \dots) / (n+1)^2 = \\
 &= [\sum_1^n (F_x(Z_i) + 1 - F_z(X_i)) + P(X \leq Z) - ng_n] / (n+1)^2.
 \end{aligned}$$

Using a known formula for variances we get

$$\begin{aligned}
 V(r_n) &= E(V(r_n | X)) + V(E(r_n | X)) = \\
 &= E(V(r_n | X)) + V(P(X \leq Z) / (n+1) + \sum_1^n (1 - F_z(X_i) - 1 + F_z(X_i)) / (n+1)^2) = \\
 &= E(V(\sum_1^n (F_x(Z_i) - F_{n,x}(Z_i)) | X)) / (n+1)^4 + 0 = \\
 &= E(V(F_x(Z_1) - F_{n,x}(Z_1) | X)) n / (n+1)^4 = \\
 &= o(n^{-3}).
 \end{aligned}$$

It remains to check formula (1.6). Some computations give

$$\begin{aligned}
 \delta_n &= \text{Cov}(r_n, g_n) = E(\text{Cov}(r_n, g_n | X)) + \text{Cov}(E(r_n | X), E(g_n | X)) = \\
 &= E(\text{Cov}(F_x(Z_1) - F_{n,x}(Z_1), F_{n,x}(Z_1) | X)) / (n+1)^2 = \\
 &= -E(V(F_x(Z_1) - F_{n,x}(Z_1) | X)) / (n+1)^2.
 \end{aligned}$$

The last equality follows since

$$\begin{aligned}
 E(\text{Cov}(F_x(Z_1) - F_{n,x}(Z_1), F_x(Z_1) | X)) &= \\
 &= E[(F_x(Z_1) - F_{n,x}(Z_1))(F_x(Z_1))] = \\
 &= E[(F_x(Z_1) - E(F_{n,x}(Z_1) | Z))(F_x(Z_1))] = \\
 &= 0.
 \end{aligned}$$

We finally compute the difference $((n+1)/n)^2 \delta_n - \delta_{n-1}$. It is easy to see that if this is $o(n^{-3})$ so is $\delta_n - \delta_{n-1}$.

$$\begin{aligned}
 & -(((n+1)/n)^2 \delta_n - \delta_{n-1}) = \\
 & = E[V(F_X(Z_1) - F_{n-1,X}(Z_1) + F_{n-1,X}(Z_1) - F_{n,X}(Z_1) \mid X) - \\
 & \quad - V(F_X(Z_1) - F_{n-1,X}(Z_1) \mid X)]/n^2 = \\
 & = E[V(F_{n-1,X}(Z_1) - F_{n,X}(Z_1) \mid X) + \\
 & \quad + 2\text{Cov}(F_X(Z_1) - F_{n-1,X}(Z_1), F_{n-1,X}(Z_1) - F_{n,X}(Z_1) \mid X)]/n^2 = \\
 & = o(n^{-3}).
 \end{aligned}$$

The last equality is trivial since $V(F_X(Z_1) - F_{n-1,X}(Z_1)) = O(1/n)$ and $|F_{n-1,X}(Z_1) - F_{n,X}(Z_1)| \leq 1/n$.

We have thus shown that we can take the jackknife of the normalized rank sum without changing the limiting distribution. From Theorem 1.4 the same result follows for any function of g_n that has a bounded second derivative in a neighbourhood of $P(X \leq Z)$. We have also shown that the Wilcoxon rank sum has a normal limiting distribution under any alternative such that $0 < P(X \leq Z) < 1$ as $n \rightarrow \infty$.

The method of this paper can thus be used to show normality, when this is difficult with other methods. We also understand why it sometimes is hard to check condition 1.2. It is often equally simple first to prove directly that the jackknife and the original statistic have the same asymptotic distribution and use Theorem 1.5.

REFERENCES

- Scott, D.J. (1973) Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. Adv. in Appl. Prob. 5 pp 119-137.
- [1] Thorburn, D.E. (1975) Some asymptotic properties of jackknife statistics, Technical report 1975:6, Dept. of Math. Statist. Univ. of Lund, Sweden.

SOME DIFFERENT JACKKNIFE
TECHNIQUES

by

Daniel Thorburn

1975:8

June 1975

SUMMARY

We consider generalized, higher order, complete and incomplete jackknives based on different group sizes. For these jackknives we discuss different properties such as: asymptotic equality with the original statistic, asymptotic normality, estimation of the variance and asymptotic efficiency.

KEY WORDS

Asymptotic distribution, asymptotic normality, efficiency, generalized jackknife, higher order jackknife, jackknife.

SOME DIFFERENT JACKKNIFE TECHNIQUES

by

Daniel Thorburn

1. INTRODUCTION

Let $X_1, X_2, \dots$ be i.i.d. random variables and let g_n be a real-valued and symmetric function of n variables. For simplicity let g_n also denote the random variable $g_n(X_1, \dots, X_n)$. If g_n is an estimate with a bias, it is often possible to reduce the bias by forming a suitable linear combination of g_n for different n . Such a combination is called a jackknife. The best-known jackknife is perhaps the arithmetic mean

$$(1.1) \quad \hat{g}_n = \frac{1}{n} \sum_{i=1}^n Y_{n,i}$$

of the "pseudo-values"

$$Y_{n,i} = ng_n - (n-1)g_{n,i}$$

where

$$g_{n,i} = g_{n-1}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n).$$

This jackknife removes a bias of type k/n .

In [1] and [2] the asymptotic distribution of $\hat{g}_n$ is studied. In this paper we give corresponding results for some other jackknives. In Section 2 we study generalized jackknives, i.e. jackknives that remove other types of bias than k/n and in Section 3 we discuss second and higher order jackknives, i.e. jackknives that remove two or more types of bias simultaneously. In the last three sections results are given for jackknives based on group size k , i.e. jackknives that use the difference in bias between g_n and g_{n-k} instead of between g_n and g_{n-1} . The value k is fixed in Section 4 and increases with n in Section 5. In Section 6 generalized jackknives based on an increasing group size are discussed.

In the remaining part of this section we reproduce some results from [1] and [2], which will be used in this paper. In [1] the distributions of g_n and $\hat{g}_n$ are compared. It is shown that they are asymptotically equal if two

sets of conditions hold. In [2] further conditions are given, which ensure asymptotic normality. These three conditions are given below. Here and in the sequel $o(\cdot)$ is taken as n tends to infinity.

CONDITION 1.1 $V(g_n) = \sigma^2/n + \varepsilon_n$ where

$$(1.2) \quad \varepsilon_n = o(n^{-1})$$

$$(1.3) \quad \varepsilon_n - \varepsilon_{n-1} = o(n^{-2})$$

$$(1.4) \quad \varepsilon_n - 2\varepsilon_{n-1} + \varepsilon_{n-2} = o(n^{-3}).$$

CONDITION 1.2 $E(g_{n+1} | X_1, \dots, X_n) - \frac{n}{n+1} g_n = r_n$ where

$r_n = r_n(X_1, X_2, \dots, X_n)$ is a random variable such that

$$(1.5) \quad V(r_n) = o(n^{-3})$$

$$(1.6) \quad \text{Cov}(g_n, r_n) = \delta_n = o(n^{-2})$$

$$(1.7) \quad \delta_n - \delta_{n-1} = o(n^{-3}).$$

CONDITION 1.3 There exist Y_i such that

$$E(Y_{n,i} - Y_i)^2 = o(1), \quad i = 1, 2, \dots$$

Two of the main results in [1] are the following theorems.

THEOREM 1.4 If Conditions 1.1 and 1.2 hold, then

$$(1.8) \quad V(Y_{n,i}) = \sigma^2 + o(1)$$

$$(1.9) \quad \text{Cov}(Y_{n,i}, Y_{n,j}) = o(n^{-1}) \quad i \neq j$$

$$(1.10) \quad V(\hat{g}_n - g_n) = o(n^{-1}).$$

THEOREM 1.5 Let there exist a number m such that

$$(1.11) \quad \sqrt{n}(E(g_n) - m) \rightarrow 0 \quad \text{and} \quad \sqrt{n}(E(\hat{g}_n) - m) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and let f be a function with a bounded second derivative in a neighbourhood of m . If Conditions 1.1 and 1.2 hold, then

$$\sqrt{n}(\widehat{f(g_n)} - f(g_n)) \xrightarrow{d} 0 \quad \text{as } n \rightarrow \infty.$$

In [2] the following results are proved:

THEOREM 1.6 If Conditions 1.1, 1.2 and 1.3 hold, the following four results are also true:

- a) The values Y_i defined in Condition 1.3 are independent
- b) $V(\hat{g}_n - \frac{1}{n} \sum_1^n Y_i) = o(n^{-1})$
- c) $\sqrt{n} (\hat{g}_n - m) \xrightarrow{d} N(0, \sigma^2)$ as $n \rightarrow \infty$
- d) $\tilde{s}_n^2 = \frac{1}{n-1} (\sum_1^n Y_{n,i}^2 - \frac{1}{n} (\sum_1^n Y_{n,i})^2) \xrightarrow{d} \sigma^2$.

2. THE GENERALIZED JACKKNIFE

Jackknives may be constructed that remove any type of bias. A good example is the generalized jackknife (Gray & Schucany 1972),

$$(2.1) \quad \tilde{g}_n = \frac{1}{n} \sum_1^n \frac{f(n-1)g_n - f(n)g_{n,i}}{f(n-1) - f(n)},$$

where f is any function such that $f(n-1) \neq f(n)$. If the bias of g_n is proportional to $f(n)$, then $\tilde{g}_n$ is unbiased. For $f(n) = k/n$, the generalized jackknife reduces to the usual jackknife $\hat{g}_n$. This jackknife has been used for instance in higher order jackknives, which are described in the next section, and in finite sampling (Durbin 1959).

a) Relation with $\hat{g}_n$

The variable $\tilde{g}_n$ has the same asymptotic properties as $\hat{g}_n$ for many different functions $f(n)$. This is seen by rewriting (2.1) as

$$(2.2) \quad \tilde{g}_n = \frac{f(n)}{(n-1)(f(n-1)-f(n))} \hat{g}_n + (1 - \frac{f(n)}{(n-1)(f(n-1)-f(n))}) g_n.$$

The coefficient $\frac{f(n)}{(n-1)(f(n-1)-f(n))}$ often converges to a constant when n tends to infinity. This occurs for instance when $f(n) = k n^{-a}$ for a positive a or when $f(n)$ can be written as a finite sum of such terms. From

Theorems 1.4 and 1.5 we know that $\hat{g}_n$ and g_n have the same asymptotic distribution, if some conditions hold. Then by (2.2) $\tilde{g}_n$ must also have the same asymptotic distribution. If also Condition 1.3 holds we know from Theorems 1.4 and 1.6.c that the limiting distribution must be normal. We have thus proved

THEOREM 2.1 If Conditions 1.1 and 1.2 hold and if

$$(2.3) \quad \limsup_{n \rightarrow \infty} \left| \frac{f(n)}{(n-1)(f(n-1)-f(n))} \right| < \infty$$

then

$$V(\tilde{g}_n - g_n) = o(n^{-1}).$$

If furthermore Condition 1.3 holds

$$\sqrt{n}(\tilde{g}_n - E(\tilde{g}_n)) \xrightarrow{d} N(0, \sigma^2).$$

A result for functions of g_n corresponding to Theorem 1.5 can also be formulated. The result is obvious.

One of the advantages of the jackknife is the fact that it provides a good estimate of the variance as is seen in Theorem 1.6.d. The jackknife technique described in this section can also be used for this purpose.

Defining

$$\tilde{Y}_{n,i} = \frac{f(n-1)g_n - f(n)g_{n,i}}{f(n-1) - f(n)},$$

we have the following equality:

$$\begin{aligned} \tilde{s}_n^2 &= \frac{1}{n-1} \sum_1^n (Y_{n,i} - \hat{g}_n)^2 = \\ &= \frac{(n-1)(f(n-1)-f(n))^2}{f^2(n)} \sum_1^n (\tilde{Y}_{n,i} - \tilde{g}_n)^2. \end{aligned}$$

Theorem 1.6.d says that this is a consistent estimate of the variance if Conditions 1.1, 1.2 and 1.3 hold.

We have thus shown that all the results for $\hat{g}_n$ given in [1] and [2] automatically hold also for $\tilde{g}_n$. We give a small generalization of this result.

PROPOSITION 2.2 Let f_1 and f_2 be two functions such that

$$\lim_{n \rightarrow \infty} \frac{f_i(n)}{(n-1)(f_i(n-1) - f_i(n))}$$

exists different from zero for $i=1$ and 2 . Let $\tilde{g}_{n,1}$ and $\tilde{g}_{n,2}$ be defined by (2.1) with f_1 and f_2 , respectively. Then

$$\sqrt{n}(\tilde{g}_{n,1} - g_n) \rightarrow 0 \Leftrightarrow \sqrt{n}(\tilde{g}_{n,2} - g_n) \rightarrow 0.$$

The convergence may be in distribution, in mean square or with probability one.

PROOF From formula (2.2) it easily follows that $\tilde{g}_{n,1} - g_n$ and $\tilde{g}_{n,2} - g_n$ are proportional to one another. The proportionality factor

$$\frac{f_1(n)(f_2(n-1) - f_2(n))}{f_2(n)(f_1(n-1) - f_1(n))}$$

has by assumption a non-zero limit. The result is now obvious. $\square$

b) A bound on the variance

In [1] an upper bound on the asymptotic variance of g_n is given in terms of $V(\hat{g}_j)$, $j \leq n$. Here we give a similar bound expressed in $V(\tilde{g}_j)$.

Reordering the terms in (2.1), we can write g_n as

$$(2.4) \quad g_n = \frac{f(n-1) - f(n)}{f(n-1)} \tilde{g}_n + \frac{f(n)}{f(n-1)} \frac{1}{n} \sum_{i=1}^n g_{n,i}.$$

The variable g_{n-1} can in the same way be expressed in $\tilde{g}_{n-1}$ and $g_{n-1,i}$.

We get

$$g_n = \frac{f(n-1)-f(n)}{f(n-1)} \tilde{g}_n + \frac{f(n-2)-f(n-1)}{f(n-2)} \frac{f(n)}{f(n-1)} \tilde{g}_{n-1} + \frac{f(n)}{f(n-2)} \bar{g}_{n-2}$$

where $\tilde{g}_i$ and $\bar{g}_i$ are the arithmetic means of all possible $\tilde{g}_i(X_{i1}, \dots, X_{ii})$ and $g_i(X_{i1}, \dots, X_{ii})$, respectively. Repeating this procedure, we finally get

$$(2.5) \quad g_n = \sum_{i=1}^n \frac{f(i-1) - f(i)}{f(i-1)} \frac{f(n)}{f(i)} \tilde{g}_i$$

where we for simplicity have assumed that g_n is defined for all n .

The variable $\tilde{g}_j$ can be regarded as a U-statistic. From the theory of U-statistics (cf. Fraser 1957 Section 6.5) we get

$$(2.6) \quad V(\tilde{g}_i) \leq (i/n) V(\tilde{g}_1).$$

Taking the variance of both sides of (2.5), we have

$$(2.7) \quad V(g_n) \leq \left(\sum_1^n \frac{f(i-1) - f(i)}{f(i-1)} \frac{f(n)}{f(i)} (i/n)^{1/2} (V(\tilde{g}_i))^{1/2} \right)^2.$$

We have now obtained an upper bound on the variance of g_n . We shall next try to invert this and get a lower bound on the variance of $\tilde{g}_n$.

THEOREM 2.3 If $V(g_n) = C_1 n^{-a} + o(n^{-a})$ and if $f(n) = C_2 n^{-b} + C_3 n^{-b-1} + o(n^{-b-1})$ for positive constants a and b such that $1 + 2b - a > 0$, then

$$(2.8) \quad \liminf_{n \rightarrow \infty} \frac{V(g_n)}{V(\tilde{g}_n)} \leq \left(\frac{2b}{1 + 2b - a} \right)^2.$$

PROOF The proof is by contradiction. If (2.8) does not hold, there must exist n_0 and

$$(2.9) \quad C' < ((1 + 2b - a)/2b)^2 C_1$$

such that $V(\tilde{g}_n) < C' n^{-a}$ for all $n \geq n_0$.

Before going on with the proof we note that the condition on $f(n)$ implies that

$$\frac{f(n-1) - f(n)}{f(n-1)} \frac{C_2}{f(n)} = (b + o(1)) n^{b-1}.$$

Using (2.7) we now get

$$\begin{aligned} V(g_n) &\leq \left(\sum_1^n (b + o(i)) i^{b-1} n^{-b} (i/n)^{1/2} (C' i^{-a})^{1/2} + K n^{-b-1/2} \right)^2 \leq \\ &\leq \left(\sum_1^n b C_1^{1/2} n^{-b-1/2} i^{b-1/2-a/2} + K' n^{-b-1/2} \right)^2 \leq \\ &\leq b^2 C_1 (b - a/2 + 1/2)^{-2} n^{-a} + o(n^{-a}) \end{aligned}$$

where $K n^{-b-1/2}$ is the contribution from those terms where (2.9) does not hold. This expression is a contradiction and the theorem follows. $\square$

3. HIGHER ORDER JACKKNIVES.

In many situations the bias consists of a sum of two or more linearly independent terms, e.g. $a/n + b/n^2$, where a and b are unknown constants. In this situation we can first apply the ordinary jackknife to remove the first term of the bias and change the second term into $-b/(n^2 - n)$. If we now apply the generalized jackknife we obtain an unbiased estimate. We can also go the other way round: using the generalized jackknife twice we first eliminate the bias b/n^2 and then the remaining bias $a/(2n-1)$. These are examples of second order jackknives.

The resulting statistic will be exactly the same independently of which part of the bias that was eliminated first. This follows trivially from the following two facts: these two jackknives are both linear functions of g_n , of the arithmetic mean $\bar{g}_{n-1}$ of all $g_{n,i}$ and of the arithmetic mean $\bar{g}_{n-2}$ of $g_{n-2}(X_{i_1}, \dots, X_{i_{n-2}})$ for all subsets of size $n-2$; for the combination $a_1 g_n + a_2 \bar{g}_{n-1} + a_3 \bar{g}_{n-2}$ to be unbiased it must hold that

$$a_1 + a_2 + a_3 = 1$$

$$a_1/n + a_2/(n-1) + a_3/(n-2) = 0$$

$$a_1/n^2 + a_2/(n-1)^2 + a_3/(n-2)^2 = 0.$$

If the bias consists of k linearly independent terms, we must clearly use the jackknife procedure k times. The same result is obtained by the methods of Schucany, Gray and Owen (1971). This type of statistic is called a higher order jackknife.

In this section we discuss the asymptotic distribution of higher order jackknives. From the previous section we see that it is sufficient to assume that all jackknives are of the usual type $\hat{g}_n$ if the limit in (2.3) exists.

In the remaining part of this section we shall discuss when the k :th order jackknife and the original statistic have the same asymptotic distribution. We first give without proof a condition which is a direct

generalization of Conditions 1.1 and 1.2, but which is too complicated to be of any practical use. After that we study functions of the arithmetic mean.

It is clear that we may take k successive jackknives without changing the limiting behaviour, if Conditions 1.1 and 1.2 hold for the $k-1$ first jackknives. A formulation of this in terms of moments of the original sequence we get the condition below. In order to state it in a comprehensible way we need some notations. Let Δ denote the difference operator i.e.

$$\Delta f(n) = f(n) - f(n-1).$$

Let r_n^i be defined by

$$r_n^0 = g_n$$

$$r_n^i = E(r_{n+1}^{i-1} | X_1, \dots, X_n) - \frac{n}{n+1} r_n^{i-1}.$$

In particular we have $r_n^1 = g_n$. It is possible to take k successive jackknives without changing the limiting behaviour of the statistic g_n if the following condition holds.

CONDITION 3.1

$$\Delta^a(V(g_n) - \sigma^2/n) = o(n^{-a}) \quad \text{for all integers } 0 \leq a \leq 2k$$

and

$$\Delta^a \text{Cov}(r_n^i, r_n^j) = o(n^{-a-i-j})$$

for all non-negative integers a, i and j such that $i + j > 0$ and $a + i + j \leq 2k$.

For $k = 1$ this condition is equivalent to Conditions 1.1 and 1.2. That Condition 3.1 gives the desired result may be proved by induction. It may also be proved directly by studying the covariance structure of variables like

$$Y_{n,(1,2, \dots, k)} = \sum_0^k (n-j)^k (-1)^k \sum g_{n-j}(X_{i_1}, \dots, X_{i_{k-j}}, X_{k+1}, \dots, X_n)$$

where the second summation is performed over all subsets of size $k-j$ of the numbers $1, 2, \dots, k$. The k :th order jackknife is then the mean of all such variables.

However, the proofs are long and complicated and the result is unimportant since Condition 3.1 is hard to check. Because of this we leave it unproved that k jackknives may be taken if Condition 3.1 holds.

Next we study functions of the arithmetic mean. We denote the k :th order jackknife of g_n by

$$\hat{\Lambda}^k(g_n).$$

This means that

$$\hat{\Lambda}^1(g_n) = \hat{g}_n \quad \text{and} \quad \widehat{\hat{\Lambda}^k(g_n)} = \hat{\Lambda}^{k+1}(g_n).$$

THEOREM 3.2 Let $V(X_1) = \sigma^2$ and let f be a function with $2k+2$ bounded derivatives in a neighbourhood of $E(X_1)$. Then

$$\sqrt{n} |\hat{\Lambda}^k(f(\bar{X})) - f(\bar{X})| \xrightarrow{d} 0.$$

PROOF The proof is divided into three parts: i) all observations can be assumed to lie in a neighbourhood of $E(X_1)$ where f has $2k+2$ bounded derivatives; ii) it is sufficient to study functions of type $f(g_n) = g_n^i$, $0 \leq i \leq 2k+1$; iii) the theorem holds for such functions.

We assume for simplicity that $E(X_1) = 0$. In the proof we will use g_n instead of $\bar{X}$ and we let $\bar{g}_{n-j}$ denote the arithmetic mean of all possible $g_{n-j}(X_{i_1}, \dots, X_{i_{n-j}})$.

i) Suppose that $f^{(2k+2)}(\cdot) \leq M$ in the interval $(-d, d)$. We know from the central limit theorem that

$$P(-d \leq g_n \leq d) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

This implies that

$$P(-d \leq g_i \leq d, \quad i = n-k, \dots, n) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

This means that the limiting probabilities are not changed if we assume that g_i lies in the interval $(-d, d)$.

ii) Since f has $2k+2$ derivatives it can be expanded into a power series,

$$f(x) = \sum_0^{2k+1} a_i x^i + f^{(2k+2)}(\xi) x^{2k+2}$$

where ξ is a number between 0 and x . Since $f^{(2k+2)}$ is bounded, we get (using i) above) that

$$|f(g_n) - \sum_0^{2k+1} a_i g_n^i| \leq M |g_n^{2k+2}|.$$

Taking k successive jackknives we get

$$\sqrt{n} |\wedge^k(f(g_n)) - \sum_0^{2k+1} a_i \wedge^k(g_n^i)| \leq \sqrt{n} M \sum_0^k \binom{k}{i} n^i (n-1)^{k-i} |\overline{g_{n-i}^{2k+2}}|.$$

Since $\sqrt{n} g_n$ has an asymptotic normal distribution, this expression tends to zero in distribution. It is thus sufficient to prove the theorem for

$$f(g_n) = g_n^i \quad i = 0, \dots, 2k+1.$$

iii) We first rewrite $\wedge^k(g_n^i)$:

$$\begin{aligned} \wedge^k(g_n^i) - \wedge^{k-1}(g_n^i) &= (n-1) \wedge^{k-1}(g_n^i - \overline{g_{n-1}^i}), \\ \wedge^{k-1}(g_n^i - \overline{g_{n-1}^i}) - \wedge^{k-2}(g_n^i - \overline{g_{n-1}^i}) &= (n-1) \wedge^{k-2}(g_n^i - 2\overline{g_{n-1}^i} + \overline{g_{n-2}^i}). \end{aligned}$$

Repeating this procedure, we finally get

$$\wedge^k(g_n^i) + \sum_1^k (-1)^j \binom{k}{j} \wedge^{k-j}(g_n^i) = (n-1)^k \sum_0^k (-1)^j \binom{k}{j} \overline{g_{n-j}^i}.$$

We shall make a proof by induction. If we assume that the theorem is true for all integers less than k , the second term on the left hand side behaves as $-g_n^i$. The theorem is proved if we can show that

$$(3.1) \quad \sqrt{n} (n-1)^k \sum_0^k (-1)^j \binom{k}{j} \overline{g_{n-j}^i} \xrightarrow{d} 0.$$

Since g_n is the arithmetic mean, this is a homogeneous function of the variables $X_1, X_2, \dots, X_n$. We shall study a typical term. Let $\alpha_1, \dots, \alpha_n$ be numbers such that $\sum \alpha_j = i$. The coefficient in front of $\prod_1^n X_j^{\alpha_j}$ is

$$\sqrt{n} (n-1)^k \sum_0^k (-1)^j \binom{k}{j} (n-j)^{-i} \binom{n-j}{v} i! / \left(\binom{n}{v} \prod_1^n \alpha_j! \right)$$

where v is the number of non-zero elements in $(\alpha_1, \dots, \alpha_n)$. Let $\overline{X}^\alpha$

denote the mean of all possible $\prod_1^n X_j^{\alpha_{ij}}$. The coefficient in front of this mean is a constant times

$$\sqrt{n} (n-1)^k \sum_0^k (-1)^j \binom{k}{j} (n-j)^{-i} \binom{n-j}{v}.$$

After some computations we find that this expression is of order

$$O(n^{1/2 + k - k - i + v}) \text{ when } v \neq i \text{ and}$$

$$O(n^{1/2 + k - k - i + v - 1}) \text{ when } v = i.$$

Since $V(X_i) = \sigma^2$ we have that

$$\overline{X}^\alpha n^{-i + v + v_2} \xrightarrow{d} 0$$

where v_2 is the number of elements larger than one in $(\alpha_1, \dots, \alpha_n)$

These two formulas together give that the $\overline{X}^\alpha$ term in (3.1) tends to zero in distribution. Since there are only a finite number of such terms, (3.1) holds.

It only remains to complete the induction by observing that the theorem is true for $k = 0$. □

This theorem generalizes a result of Millér (1964). He proved a somewhat stronger result in the case $k = 1$. In our theorem the assumption of $2k+2$ derivatives can probably be weakened.

We proved the result for k ordinary successive jackknives. It is obvious that ordinary can be replaced by generalized if (2.3) holds:

COROLLARY 3.3 Let $V(X_i)$ exist and let f be a function with $2k+2$ bounded derivatives in a neighbourhood of $E(X_i)$. Let $\sim^k$ denote a k :th order generalized jackknife removing k types of bias which all obey (2.3). Then

$$\sqrt{n} | \sim^k(f(\bar{X})) - f(\bar{X}) | \xrightarrow{d} 0.$$

4. JACKKNIVES BASED ON GROUP SIZE k

The jackknife $\hat{g}_n$ is a linear combination of a statistic g_n calculated from the whole sample and n statistics $g_{n,i}$ based on all variables but one. If we instead combine g_n with statistics based on $n-k$ variables, we get a "jackknife based on group size k ".

For example we may take

$$(4.1) \quad \hat{g}_n^{(k)} = \frac{1}{k \binom{n}{k}} \sum (ng_n - (n-k)g_{n-k}(X_{i_1}, \dots, X_{i_{n-k}}))$$

where the summation is performed over all subsets $i_1, \dots, i_{n-k}$ of size $n-k$. Or assuming that $p = n/k$ is an integer we may take

$$(4.2) \quad \hat{g}_{n(k)} = \frac{1}{p} \sum_1^p (pg_n - (p-1)g_{n-k}(X_1, \dots, X_{ik-k}, X_{ik+1}, \dots, X_n)).$$

a) Complete jackknives

The complete jackknife (4.1) may be rewritten

$$\begin{aligned} (4.3) \quad \hat{g}_n^{(k)} &= \frac{1}{k} \left[\frac{n}{\binom{n}{0}} g_n - \frac{n-k}{\binom{n}{k}} \sum g_{n-k}(X_{i_1}, \dots, X_{i_{n-k}}) \right] = \\ &= \frac{1}{k} \sum_{j=0}^{k-1} \left[\frac{n-j}{\binom{n}{j}} \sum g_{n-j}(X_{i_1}, \dots, X_{i_{n-j}}) - \right. \\ &\quad \left. - \frac{n-j-1}{\binom{n}{j+1}} \sum g_{n-j-1}(X_{i_1}, \dots, X_{i_{n-j-1}}) \right] = \\ &= \frac{1}{k} \sum_{j=0}^{k-1} \frac{1}{\binom{n}{j}} \sum \left[(n-j)g_{n-j}(X_{i_1}, \dots, X_{i_{n-j}}) - \right. \\ &\quad \left. - (n-j-1) \sum_1^{n-j} g_{n-j,i}(X_{i_1}, \dots, X_{i_{n-j}}) \right] = \\ &= \frac{1}{k} \sum_{j=0}^{k-1} \frac{1}{\binom{n}{j}} \sum \hat{g}_{n-j}^{(k)}(X_{i_1}, \dots, X_{i_{n-j}}), \end{aligned}$$

where all the inner sums are performed over all subsets of the corresponding size, without respect to order. If we instead sum over all permutations of the first n integers, (4.3) equals

$$\frac{1}{k} \sum_{j=0}^{k-1} \frac{1}{n!} \sum \hat{g}_{n-j}^{(k)}(X_{i_1}, \dots, X_{i_{n-j}}).$$

A change in the order of summation gives

$$(4.4) \quad \hat{g}_n^{(k)} = \frac{1}{n!} \sum \frac{1}{k} \sum_{j=0}^{k-1} \hat{g}_{n-j}^{(k)}(X_{i_1}, \dots, X_{i_{n-j}}).$$

We now prove a result that corresponds to formula (1.10) and Theorem 1.6.d in this situation.

THEOREM 4.1 If Conditions 1.1 and 1.2 hold

$$V(\hat{g}_n^{(k)} - g_n) = o(n^{-1}).$$

If furthermore Condition 1.3 holds

$$\sqrt{n}(\hat{g}_n^{(k)} - E(\hat{g}_n^{(k)})) \xrightarrow{d} N(0, \sigma^2).$$

PROOF From Conditions 1.1 and 1.2 it is easy to prove that

$$V(g_n - g_{n-j}) = o(n^{-1}) \quad \text{for all } j \leq k.$$

From Theorem 1.4 we have that

$$V(\hat{g}_{n-j}^{(k)} - g_{n-j}) = o(n^{-1}) \quad \text{for all } j \leq k.$$

These two results together imply that

$$V\left(\frac{1}{k} \sum_{j=0}^{k-1} \hat{g}_{n-j}^{(k)} - g_n\right) = o(n^{-1}).$$

By (4.4) $\hat{g}_n^{(k)} - g_n$ is an arithmetic mean of such variables. We have thus proved the first part of the theorem. The second part follows directly from Theorem 1.6. □

It is also possible to construct an estimate of the variance based on these differences,

$$\frac{n}{k(n-1)\binom{n}{k}} \sum (ng_n - (n-k)g_{n-k}(X_{i_1}, \dots, X_{i_{n-k}}) - \hat{g}_n^{(k)})^2$$

which can be shown to be consistent (e.g. using Theorem 4.2 below).

b) Incomplete jackknives

Since the calculation of a complete jackknife often requires large computations, an incomplete jackknife such as (4.2) is more convenient to

use. It is obvious from the Rao-Blackwell theorem that this version cannot be better than the complete one. However, the asymptotic efficiency is the same as we shall now show.

The incomplete jackknife (4.2) can be considered as an ordinary jackknife (1.1), if the k -tuple $X_{ik-k+1}, \dots, X_{ik}$ is regarded as the i :th observation. All the results for the ordinary jackknife in the papers [1] and [2] thus continue to hold also for the jackknife based on step size k , if Conditions 1.1 and 1.2 are assumed to hold for the k -tuples instead of just for $k=1$. We shall show that a sufficient condition for this is that the same conditions hold only for $k=1$. Indeed, if this assumption is true it follows easily that Condition 1.1 holds for the k -tuples. To show that also Condition 1.2 holds we note that

$$\begin{aligned} E(g_{n+k} | X_1, \dots, X_n) - \frac{p}{p+1} g_n &= \\ &= E(r_{n+k-1} + r_{n+k-2} + \dots + r_n | X_1, \dots, X_n), \end{aligned}$$

where r_n is defined in Condition 1.2. The variance of this expression is bounded by

$$V(r_{n+k-1} + \dots + r_n) \leq k^2 o(n^{-3})$$

It remains to check formula (1.7),

$$\begin{aligned} \text{Cov}(g_n, E(r_{n+k-1} + \dots + r_n | X_1, \dots, X_n)) - \\ - \text{Cov}(g_{n-k}, E(r_{n-1} + \dots + r_{n-k} | X_1, \dots, X_{n-k})) = \\ = \sum_1^k (\text{Cov}(g_n, E(r_{n+k-i} | X_1, \dots, X_n)) - \\ - \text{Cov}(g_{n-k}, E(r_{n-i} | X_1, \dots, X_{n-k}))) = \\ = \sum_1^k (\text{Cov}(g_n, r_{n+k-i}) - \text{Cov}(g_{n-k}, r_{n-i})). \end{aligned}$$

Here the last equality follows from the independence of g_n and

$X_{n+1}, X_{n+2}, \dots$. We shall now prove that each term in this sum is of order $o(n^{-3})$:

$$\begin{aligned}
& \text{Cov}(g_n, r_{n+k-i}) - \text{Cov}(g_{n-k}, r_{n-i}) = \\
& = \frac{1}{n-k} \text{Cov}((n-i)g_{n-i} - (n-k)g_{n-k}, r_{n-i}) - \frac{n-i}{n-k} \delta_{n-i} - \\
& \quad - \frac{1}{n} \text{Cov}((n+k-i)g_{n+k-i} - ng_n, r_{n+k-i}) + \frac{n+k-i}{n} \delta_{n+k-i} = \\
& = \sum_{j=i}^{k-1} \left(\frac{1}{n-k} \text{Cov}((n-j)g_{n-j} - (n-j-1)g_{n-j-1}, r_{n-i}) - \right. \\
& \quad \left. - \frac{1}{n} \text{Cov}((n+k-j)g_{n+k-j} - (n+k-j-1)g_{n+k-j-1}, r_{n+k-i}) \right) + o(n^{-3}) = \\
(4.5) \quad & = \sum_{j=i}^{k-1} \left(\frac{1}{n-k} \text{Cov}(\hat{g}_{n-j}, r_{n-i}) - \frac{1}{n} \text{Cov}(\hat{g}_{n+k-j}, r_{n+k-i}) \right) + o(n^{-3}).
\end{aligned}$$

The first of these equalities is an addition and subtraction of the δ -terms. In the second equality the covariances are expanded into telescoping sums. For the last equality we use that the covariances with pseudo-values are the same as the covariances with their arithmetic mean, the jackknife. This follows from the symmetry. We know the variance of $\hat{g}_{n-j}$ from Theorem 1.4 and that of r_{n-i} from Condition 1.2. Together these two results give that the covariances in (4.5) are $o(n^{-2})$. The whole expression (4.5) is thus of order $o(n^{-3})$ and hence we have proved:

THEOREM 4.2 The results of Theorems 1.4, 1.5 and 1.6 hold also for the incomplete jackknife $\hat{g}_{n(k)}$.

5. JACKKNIVES BASED ON AN INCREASING GROUP SIZE

The group size k in the jackknives (4.1) and (4.2) may also increase with n . In this section we assume in the first part that $p = n/k$ is a fixed integer larger than one. In the second part we let n/k tend to one. In the last part we introduce a new incomplete jackknife.

a) n/k fixed

We shall study the complete and the incomplete jackknives simultaneously. From a practical point of view only the latter is of interest, since the first one involves too much computational work.

When $p=2$ the incomplete jackknife was originally proposed by Quenouille (1949). Brillinger (1966) studied a little more general situation. He used also conditions on the moments, but his conditions are different from ours.

We first prove a result corresponding to (1.10).

THEOREM 5.1 If Conditions 1.1, 1.2 and 1.3 hold and $p = n/k$ is a fixed integer, then

$$(5.1) \quad V(\hat{g}_n^{(n/p)} - g_n) = o(n^{-1})$$

and

$$(5.2) \quad V(\hat{g}_{n(n/p)} - g_n) = o(n^{-1}).$$

PROOF We start by proving (5.2). By definition (4.2)

$$(5.3) \quad \hat{g}_{n(n/p)} = \frac{1}{p} \sum_1^p (pg_n - (p-1)g_{n-k}(X_1, \dots, X_{ik-k}, X_{ik+1}, \dots, X_n))$$

From Theorems 1.4 and 1.6 we have

$$(5.4) \quad V(g_n - \frac{1}{n} \sum_1^n Y_i) \leq 2V(g_n - \hat{g}_n) + 2V(\hat{g}_n - \frac{1}{n} \sum_1^n Y_i) = o(n^{-1}).$$

Using this in (5.3) we get

$$\begin{aligned} V(\hat{g}_{n(n/p)} - g_n) &= \\ &= V\left[\left(\frac{1}{n} \sum_1^n Y_i - g_n\right) + \frac{1}{p} \sum_1^p \left(p(g_n - \frac{1}{n} \sum_1^n Y_j) - \right.\right. \\ &\quad \left.\left. - (p-1)(g_{n-k}(X_1, \dots, X_{ik-k}, X_{ik+1}, \dots, X_n) - \frac{1}{n-k} \sum_{\substack{1 \leq j \leq ik-k \\ ik+1 \leq j \leq n}} Y_j)\right)\right] = \\ &= o(n^{-1}), \end{aligned}$$

since this is the variance of a finite number of terms, all of which tend to zero in mean square by (5.4).

The result (5.1) now follows, since (5.1) is the arithmetic mean of a number of terms distributed as (5.2). □

The variance estimator that corresponds to $\hat{g}_{n(n/p)}$ is

$$(5.5) \quad t_n^2 = \frac{n}{(p-1)p} \left(\sum_1^p (pg_n - (p-1)g_{n-k}(X_1, \dots, X_{ik-k}, X_{ik+1}, \dots, X_n) - \hat{g}_{n(n/p)})^2 \right).$$

This estimator t_n^2 has an asymptotic χ^2 -distribution after a suitable normalization.

THEOREM 5.2 If Conditions 1.1, 1.2 and 1.3 hold and if p is a fixed integer, then

$$(p-1) t_n^2 / \sigma^2 \xrightarrow{d} \chi^2(p-1) \text{ as } n \rightarrow \infty$$

$$\sqrt{n} (\hat{g}_{n(n/p)} - E(\hat{g}_{n(n/p)})) / t_n \xrightarrow{d} t(p-1) \text{ as } n \rightarrow \infty$$

where $t(p-1)$ denotes the studentized t -distribution with $p-1$ degrees of freedom.

PROOF From (5.4) we have

$$(5.6) \quad V\left(\frac{1}{\sqrt{n}} (ng_n - \sum_1^n Y_i - (n-k)g_{n-k} - \sum_1^{n-k} Y_i)\right) = o(1).$$

Using Theorem 1.6.a we get from the central limit theorem that

$$(5.7) \quad \frac{1}{\sqrt{k}} \sum_{ik+1}^{ik+k} (Y_j - E(Y_j)) \xrightarrow{d} N(0, \sigma^2) \text{ as } n \rightarrow \infty$$

and that the values on the left hand side are independent for different i . Combining formulas (5.6) and (5.7) we obtain the theorem. $\square$

Theorem 5.2 is the same result as that proved by Brillinger (1964) in another situation.

b) n/k tends to one

We shall also discuss what happens with the complete jackknife (4.1), when n/k tends to one as n tends to infinity. For the incomplete jackknife this is impossible since $p = n/k$ must be an integer for the definition (4.2) to make sense.

Of course, we must still assume that $n-k$ tends to infinity. Else we cannot say anything about the bias or the distribution. The statistic $\hat{g}_n^{(k)}$ always removes a bias c/n completely. However, a bias of smaller order $c n^{-a}$, say, is enlarged to $c n^{-1} (n-k)^{1-a}$ if $a > 1$.

We shall next show that $V(\hat{g}_n^{(k)} - g_n) = o(V(g_n))$ under fairly general conditions. Reordering the terms in (4.1) we get

$$(5.8) \quad \hat{g}_n^{(k)} = \frac{n-k}{k} \left[\frac{1}{\binom{n}{k}} \sum g_{n-k}(X_{i_1}, \dots, X_{i_{n-k}}) - g_n \right]$$

where the summation is over all subsets. The sum in this expression may be regarded as a U-statistic. From the theory of U-statistics (cf. Fraser 1957 Section 6.5) we have the following result

$$V\left(\frac{1}{\binom{n}{k}} \sum g_{n-k}(X_{i_1}, \dots, X_{i_{n-k}})\right) \leq \frac{n-k}{n} V(g_{n-k}).$$

The variance of the second term of (5.8) is thus at most

$$(5.9) \quad 2\left(\frac{n-k}{k}\right)^2 \left(\frac{n-k}{n} V(g_{n-k}) + V(g_n)\right).$$

A comparison of this with the variance of $\hat{g}_n$ gives the following theorem.

THEOREM 5.3 Let $V(g_n) = C n^{-a} + o(n^{-a})$ for some positive constant C and some constant a , $0 \leq a < 3$, and let this variance exist. If $k/n \rightarrow 1$, then

$$V(\hat{g}_n^{(k)} - g_n) = o(V(g_n)).$$

PROOF We can always find a C' such that $V(g_n) \leq C'(n^{-a})$ for all $n > 1$.

In this situation we have

$$2\left(\frac{n-k}{k}\right)^2 \left(\frac{n-k}{n} V(g_{n-k})\right) \leq 2C' \left(1 - \frac{k}{n}\right)^{3-a} n^{-a} + o\left(\left(1 - \frac{k}{n}\right)^{3-a} n^{-a}\right).$$

This is of a smaller order than the variance of $\hat{g}_n$. We thus get that

(5.9) is of a smaller order than the variance of $\hat{g}_n$. The result now follows from (5.8).

This theorem applies to much more general situations than those discussed earlier in this paper, e.g. to the minimum of a uniform distribution.

A complete jackknife of this type, applied to the minimum, will by the theorem have the same asymptotic distribution as the minimum. The bias will be of order $1/(n(n-k))$. The complete jackknives are seldom easy to compute but the jackknife of the minimum can be simplified:

$$\hat{X}_{(1,n)}^{(k)} = X_{(1,n)} + \sum_1^k \binom{n-k}{1} \frac{(k-1)!(n-i)!}{(k-i)!n!} (X_{(i,n)} - X_{(i+1,n)})$$

where $X_{(i,n)}$ is the i :th order statistic among the first n observations.

c) An incomplete jackknife when $n/(n-k)$ is an integer

Another kind of jackknife is obtained in the following way. Suppose that $n = (n-k)q$ where q is an integer. Divide the sample into q groups with $n-k$ elements each. Let

$$g_{i:n} = g_{n-k}(X_{i(q-1)+1}, \dots, X_{iq}).$$

Define the jackknife by

$$\check{g}_n = \frac{q}{q-1} g_n - \frac{1}{q(q-1)} \sum_1^q g_{i:n}.$$

This statistic can be regarded as an incomplete version of $\hat{g}_n^{(k)}$, defined in (4.1), which is the arithmetic mean of $\check{g}_n$ for all possible divisions into q groups. The mean value properties of $\check{g}_n$ are thus the same as those of $\hat{g}_n^{(k)}$.

In the same way as in the previous sections we immediately get a lot of results.

PROPOSITION 5.4 If Conditions 1.1, 1.2 and 1.3 hold and if q is a fixed integer, then

$$V(\check{g}_n - g_n) = o(n^{-1}).$$

PROOF Using (5.4) we have:

$$(5.10) \quad V(g_n - \frac{1}{n} \sum_1^n Y_i) = o(n^{-1})$$

$$(5.11) \quad V(g_{i:n} - \frac{1}{n-k} \sum_1^{n-k} Y_{i(q-1)+j}) = o(n^{-1}).$$

These two formulas together prove the proposition. $\square$

PROPOSITION 5.5 If q tends to infinity and if $V(g_n) = C n^{-a} + o(n^{-a})$ ($0 \leq a < 3$), then

$$V(\bar{g}_n - g_n) = o(V(g_n)).$$

PROOF We have that

$$\begin{aligned} V(\bar{g}_n - g_n) &= V\left(\frac{1}{q-1} g_n - \frac{1}{q(q-1)} \sum_{i=1}^q g_{i:n}\right) \leq \\ &\leq \frac{2}{(q-1)^2} V(g_n) + \frac{2}{q(q-1)^2} V(g_{i:n}) = \\ &= o(V(g_n)) + 2/(q(q-1)^2) (C(q/n)^a - o((q/n)^a)) = o(V(g_n)). \end{aligned}$$

□

Using this proposition and the fact that $\hat{g}_n^{(k)}$ is an average of $\bar{g}_n$, we have another way of proving Theorem 5.3, when q is an integer.

We end this section with a discussion on how to estimate the variance in this situation. This is quite simple since the variables $g_{i:n}$ are independent with a variance $\sigma^2/(n-k) + o(n^{-1})$, if q is fixed.

PROPOSITION 5.6 If $V(g_n) = \sigma^2/n + o(n^{-1})$ and $n-k$ increases with n , then

$$s_n^2 = \frac{n-k}{q-1} \left(\sum_{i=1}^q g_{i:n}^2 - \frac{1}{q} \left(\sum_{i=1}^q g_{i:n} \right)^2 \right)$$

is an estimate of σ^2 . If Conditions 1.1, 1.2 and 1.3 hold, the estimate is consistent when q tends to infinity and has an asymptotic χ^2 -distribution with $q-1$ degrees of freedom when q is fixed.

PROOF The variables $g_{i:n}$ are independent with a variance $\sigma^2/(n-k) + o(\frac{1}{n-k})$. From this and Condition 1.3 we get that s_n^2 is consistent if q tends to infinity.

Conditions 1.1, 1.2 and 1.3 ensures that $g_{i:n}$ converges in distribution to something with variance σ^2 . By Theorem 1.6 this limiting distribution is normal and the last statement of the proposition follows.

□

6. GENERALIZED JACKKNIVES BASED ON GROUP SIZE k

In Section 2 we discussed the generalized jackknife based on group size one, $\tilde{g}_n$. A generalized jackknife may also be based on larger group sizes. We shall in this section only discuss complete statistics, but the results are similar for the incomplete versions.

a) Asymptotic distribution

The main tool in Section 2 was that $\tilde{g}_n$ may be written as a linear combination of $\hat{g}_n$ and g_n (cf. formula (2.2)). The same thing is true for generalized jackknives based on larger group sizes.

$$(6.1) \quad \tilde{g}_n^{(k)} = \frac{1}{\binom{n}{k}} \sum \frac{f(n-k)g_n - f(n)g_{n-k}(X_{i1}, \dots, X_{in-k})}{f(n-k) - f(n)} =$$

$$\frac{f(n)}{f(n-k) - f(n)} \frac{k}{n-k} \hat{g}_n^{(k)} + \left(1 - \frac{f(n)}{f(n-k) - f(n)} \frac{k}{n-k}\right) g_n.$$

In Sections 4 and 5 we gave conditions for $\hat{g}_n^{(k)}$ and g_n to have the same asymptotic distribution. It is thus interesting to know when the coefficient before $\hat{g}_n^{(k)}$ in (6.1) tends to a constant different from plus or minus infinity. After some calculations we get the following answers for different behaviours of k :

- i) k fixed. The limit exists if $f(n) = c n^{-a} + o(n^{-a-1})$ where a is positive or if $f(n)$ is a finite sum of such terms.
- ii) k/n fixed. The limit exists if $f(n) = c n^{-a} + o(n^{-a})$ for some positive a .
- iii) $k/n \rightarrow 1$. The limit exists if $f(n) = c n^{-a} + o(n^{-a})$ for some $a \geq 1$.

Without any assumptions on the behaviour of k the following result can be proved:

PROPOSITION 6.1 If $\sqrt{n}(\hat{g}_n^{(k)} - g_n) \rightarrow 0$ and if $f(n) = \sum_1^\infty c_i n^{-i}$, then

$$\sqrt{n}(\tilde{g}_n^{(k)} - g_n) \rightarrow 0.$$

The convergence may be in distribution, in mean square or almost sure.

b) A bound on the variance

In this subsection we generalize the result of Theorem 2.3. When k is fixed it is obvious that the same result holds for $\tilde{g}_n^{(k)}$ as for $\tilde{g}_n$, but when $(n-k)/n = q$ is a fixed number the situation is much more complicated.

Rewriting the definition we get

$$(6.2) \quad g_n = \frac{f(n-k) - f(n)}{f(n-k)} \tilde{g}_n^{(k)} - \frac{1}{\binom{n}{k}} \sum \frac{f(n)}{f(n-k)} g_{qn}(X_{i_1}, \dots, X_{i_{qn}}).$$

In Section 2 we could express $\tilde{g}_{n-1}$ in terms of $\tilde{g}_{n-1}$ and $\tilde{g}_{n-2}$. Here we do not know that $q^2 n$ is an integer, or in other words we cannot be certain that g_{qn} can be expressed in $\tilde{g}_{qn}$ and g_{q2n} . If we for the moment forget this problem and assume that $nq^m = 1$, for a certain number m , we shall, in the same way as formula (2.5) get that

$$g_n = \sum_1^m \frac{f(q^i n) - f(q^{i-1} n)}{f(q^i n)} \frac{f(n)}{f(q^{i-1} n)} \tilde{g}_{q^{i-1} n}^{(\cdot)}$$

where the point stands for $((1-q)q^{i-1}n)$. From this formula bounds can be obtained in the same manner as in Section 2, e.g.

$$(6.3) \quad V(g_n) \leq \left(\sum_1^m \frac{f(q^i n) - f(q^{i-1} n)}{f(q^i n)} \frac{f(n)}{f(q^{i-1} n)} q^{(i-1)/2} (V(\tilde{g}_{q^{i-1} n}^{(\cdot)}))^{1/2} \right)^2.$$

The assumption $nq^m = 1$ implies for instance that $1/q$ is an integer and it is thus rather restrictive.

Suppose that we want $\tilde{g}_n^{(k)}$ to be a certain preassigned sequence. In this paragraph we examine how much g_n can vary, if the sequence $\tilde{g}_n^{(k)}$ is fixed. Let $q = \mu/\nu$, where μ and ν are integers with no common divisor. When ν is not a divisor of n , the variable g_n can be arbitrarily chosen. We then use (6.2) to define the remaining g_n . If we know g_n for a certain n , we by (6.2) also know as many new g_n as the number of times n can be divided by μ . If $\mu > 1$, we can choose g_n so that $V(g_n)$ and $V(g_{q^{-i}n})$, for all possible i , are larger than any desired value. A bound like (6.3) can thus be obtained only when $\mu = 1$.

A bound like that in Theorem 2.3 can, however, be obtained, since there we make an assumption on the smoothness of $V(g_n)$.

THEOREM 6.2 If $V(g_n) = c_1 n^{-a} + o(n^{-a})$ and $f(n) = c_2 n^{-b} + o(n^{-b})$ for positive constants a and b such that $1 + 2b - a > 0$, then

$$(6.4) \quad \liminf_{n \rightarrow \infty} \frac{V(g_n)}{V(\tilde{g}_n^{(k)})} \leq \left(\frac{1 - q^b}{1 - q^{b + 1/2 - a/2}} \right)^2.$$

PROOF The proof is by contradiction. If (6.4) does not hold there must exist n_0 and

$$C' < \left(\frac{1 - q^{b + 1/2 - a/2}}{1 - q^b} \right)^2 \cdot c_1$$

such that

$$V(\tilde{g}_n^{(\cdot)}) < C' n^{-a}, \quad \forall n > n_0.$$

We observe that given $\varepsilon > 0$ there exists an $n_1 > n_0$ such that

$$\left| \frac{f(qn) - f(n)}{f(qn) f(n)} c_2 n^{-b} - (1 - q^b) \right| \leq \varepsilon \quad \forall n > n_1,$$

$$|f(n)n^{-b}/c_2 - 1| \leq \varepsilon \quad \forall n > n_1 \quad \text{and}$$

$$|V(g_n)n^a - c_1| \leq \varepsilon \quad \forall n > n_1.$$

Suppose that $q = \mu/\nu$. Let m be so large that

$$q^{m(1+2b-a)/2} < \frac{\varepsilon(1-\varepsilon)}{(1+\varepsilon)^2} C'.$$

Supposing that n is a multiple of ν^m and that $q^m n$ is larger than n_1 , we have

$$g_n = \sum_1^m \frac{f(q^i n) - f(q^{i-1} n)}{f(q^i n)} \frac{f(n)}{f(q^{i-1} n)} \bar{g}_{q^{i-1} n} + \frac{f(n)}{f(q^m n)} \bar{g}_{q^m n}$$

where $\bar{g}_{q^m n}$ is the mean of all possible $g_{q^m n}(X_{i_1}, \dots, X_{i_{q^m n}})$. This is obtained from (6.2) in the same way as (2.5) is obtained from (2.4).

Taking the variance of both sides we get (using formula (2.6))

$$\begin{aligned}
V(g_n) &\leq \left[\sum_1^m \frac{f(q^i n) - f(q^{i-1} n)}{f(q^i n)} \frac{f(n)}{f(q^{i-1} n)} q^{(i-1)/2} (V(\tilde{g}_{q^{i-1} n}))^{1/2} + \right. \\
&\quad \left. + q^{m+b} \left(\frac{1+\varepsilon}{1-\varepsilon} \right) q^{m/2} (V(g_{q^m n}))^{1/2} \right]^2 \leq \\
&\leq \left[\sum_1^\infty (1-q^{b+\varepsilon})(1+\varepsilon) q^{1b-b} q^{(i-1)/2} C^{1/2} q^{-(i-1)a/2} n^{-a/2} + \right. \\
&\quad \left. + q^{(b+1/2)m} C_1^{1/2} q^{-ma/2} n^{-a/2} (1+\varepsilon)^2 / (1-\varepsilon) \right]^2 \leq \\
&\leq [(1-q^{b+\varepsilon})(1+\varepsilon) C^{1/2} n^{-a/2} / (1-q^{(1-a+2b)/2}) + \varepsilon n^{-a/2}]^2
\end{aligned}$$

By letting ε tend to zero we obtain a contradiction. $\square$

As a corollary we get a corresponding result for $g_n^{(k)}$.

COROLLARY 6.3 If $V(g_n) = c_1 n^{-a} + o(n^{-a})$; $0 < a < 3$,

then

$$\liminf_{n \rightarrow \infty} \frac{V(g_n)}{V(g_n^{\wedge(k)})} \leq \left(\frac{1-q}{1-q^{(3-a)/2}} \right)^2.$$

It is interesting to note that if we let q tend to one, the bound in Theorem 6.2 is changed into the bound of Theorem 2.3. Letting q tend to zero we get half the result of Theorem 5.3.

If we suppose that not only $V(g_n)$ but also $V(g_n^{\wedge(k)})$ essentially is a power of n , the limit inferior in Theorem 6.2 is replaced by a normal limit. In the special case $a = 1$ we have the following corollary:

COROLLARY 6.4 If $V(g_n) = \sigma^2/n + o(1/n)$ and $f(n) = cn^{-b} + o(n^{-b})$ for some positive b , then neither $V(g_n^{\wedge(k)})$, $V(\tilde{g}_n)$, $V(g_n^{\wedge(k)})$ nor $V(\tilde{g}_n^{(k)})$ can be of a smaller order than $V(g_n)$, e.g.

$$\lim_{n \rightarrow \infty} V(\tilde{g}_n^{(k)}) / V(g_n) \geq 1.$$

REFERENCES

- Brillinger, D.R. (1964) The asymptotic behaviour of Tukey's general method of setting approximate confidence limits (the jackknife) when applied to maximum likelihood estimates. *Rev. Inst. Int. Statist.* 32, pp 202-206.
- Brillinger, D.R. (1966) The applications of the jackknife to the analysis of sample surveys. *Commentary* 8, pp 74-80.
- Durbin, J. (1959) A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika* 46, pp 477-480.
- Fraser, D.A.S. (1957) *Nonparametric methods in statistics*. Wiley, New York.
- Gray, H.L. & Schucany, W.R. (1972) *The generalized jackknife statistic*. Marcel & Dekker, New York.
- Miller, R.G. (1964) A trustworthy jackknife. *Ann. Math. Statist.* 35, pp 1594-1605.
- Quenouille, M.H. (1949) Approximate tests of correlation in time series. *J. R. Statist. Soc. B* 11, pp 68-84.
- Schucany, W.R., Gray, H.L. & Owen, D.B. (1971) On bias reduction in estimation. *J. Am. Statist. Assoc.* 66, pp 524-533.
- [1] Thorburn, D.E. (1975) Some asymptotic properties of jackknife statistics. Technical report 1975:6 Dept. of Math. Statist. Univ. of Lund, Sweden.
- [2] Thorburn, D.E. (1975) On the asymptotic normality of the jackknife. Technical report 1975:7 Dept. of Math. Statist. Univ. of Lund, Sweden.

